

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
23 August 2001 (23.08.2001)

PCT

(10) International Publication Number  
**WO 01/60991 A2**

(51) International Patent Classification<sup>7</sup>: **C12N 9/12, 15/54, 5/10, A01K 67/027, C07K 16/40, C12Q 1/68, C12N 15/11, A61K 38/45, G01N 33/53**

(21) International Application Number: PCT/US01/05240

(22) International Filing Date: 16 February 2001 (16.02.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/183,682	17 February 2000 (17.02.2000)	US
60/186,559	2 March 2000 (02.03.2000)	US
60/188,606	9 March 2000 (09.03.2000)	US
60/189,998	17 March 2000 (17.03.2000)	US
60/193,851	30 March 2000 (30.03.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/183,682 (CIP)
Filed on	17 February 2000 (17.02.2000)
US	60/186,559 (CIP)
Filed on	2 March 2000 (02.03.2000)
US	60/188,606 (CIP)
Filed on	9 March 2000 (09.03.2000)
US	60/189,998 (CIP)
Filed on	17 March 2000 (17.03.2000)
US	60/193,851 (CIP)
Filed on	30 March 2000 (30.03.2000)

(71) Applicant (*for all designated States except US*): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **TANG, Y., Tom** [CN/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). **BUFORD, Neil** [GB/US]; 105 Wildwood Circle, Durham, CT 06422 (US). **GANDHI, Ameena, R.** [US/US]; 837 Roble Avenue #1, Menlo Park, CA 94025 (US). **PATTERSON, Chandra** [US/US]; 490 Sherwood Way #1, Menlo Park, CA 94024 (US). **KHAN, Farrah, A.** [IN/US]; 333 Escuela Avenue #221, Mountain View, CA 94040 (US). **YUE, Henry** [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). **HAFALIA, April** [US/US]; 2227 Calle de Primavera, Santa Clara, CA 95054 (US). **SHIH, Leo, L.**

[US/US]; Apartment B, 1081 Tanland Drive, Palo Alto, CA 94303 (US). **TRIBOULEY, Catherine, M.** [FR/US]; 1121 Tennessee Street #5, San Francisco, CA 94107 (US). **YAO, Monique, G.** [US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). **BURRILL, John, D.** [US/US]; 650 Emerald Hill Road, Redwood City, CA 94061 (US). **MARCUS, Gregory, A.** [US/US]; 1713 Connecticut Drive, Redwood City, CA 94061 (US). **ZINGLER, Kurt, A.** [US/US]; 723 Ashbury Street #C, San Francisco, CA 94117 (US). **LU, Dyung, Aina, M.** [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). **BANDMAN, Olga** [US/US]; 366 Anna Avenue, Mountain View, CA 94043 (US). **POLICKY, Jennifer, L.** [US/US]; 4864 National Avenue, San Jose, CA 95124 (US). **GRIFFIN, Jennifer, A.** [US/US]; 181 Irene Court #11, Belmont, CA 94002 (US). **THORNTON, Michael** [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). **NGUYEN, Danniel, B.** [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). **LAL, Preeti** [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). **WALSH, Roderick, T.** [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB).

(74) Agents: **HAMLET-COX, Diana et al.**; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Published:**

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: HUMAN KINASES

(57) Abstract: The invention provides human kinases (PKIN) and polynucleotides which identify and encode PKIN. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of PKIN.

WO 01/60991 A2

## HUMAN KINASES

### TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of human kinases and to the  
5 use of these sequences in the diagnosis, treatment, and prevention of cancer, immune disorders,  
disorders affecting growth and development, atherosclerosis and other cardiovascular diseases, and  
lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of  
nucleic acid and amino acid sequences of human kinases.

10

### BACKGROUND OF THE INVENTION

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of  
15 molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response  
20 to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such  
25 as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

30 There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing  
35 specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic

domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains 5 specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

10 In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) *The Protein Kinase Facts Book*, Vol I, pp. 17-20 Academic Press, San Diego CA.). In particular, two protein kinase signature sequences have been identified in the 15 kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

### Protein Tyrosine Kinases

20 Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, 25 platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and 30 prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. 35 Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in

controlling some types of cancer.

#### Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism, and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a

highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, 5  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . Fish et al., identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

10 The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the Drosophila circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

15 Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al., have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian 20 rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) Science 288:483-491).

#### Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and 25 CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the 30 cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et.al. (1995) EMBO J. 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation 35 of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) BioEssays 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be

enriched in mammalian forebrain. This protein is associated with vesicles in both axons and dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

5    Mitogen-Activated Protein Kinases

The mitogen-activated protein kinases (MAP) which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A.

- 10   Weinberg (1993) Nature 365:781-783). MAP kinase signaling pathways are present in mammalian cells as well as in yeast. The extracellular stimuli which activate MAP kinase pathways include epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, endotoxic lipopolysaccharide (LPS), and pro-inflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions  
15   including cancer, inflammation, immune disorders, and disorders affecting growth and development.

Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

- Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

- In the process of cell division, the order and timing of cell cycle transitions are under control  
30   of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role  
35   in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in

yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) *Science* 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) *Science* 277:1501-1505). Thus, activation of Chk1 prevents the damaged cell from entering mitosis. A similar deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

#### Proliferation-Related Kinases

10 Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) *J. Biol. Chem.* 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue 15 leads to oncogenic transformation.

#### 5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) *J. Biol Chem.* 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and 20 hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses, such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests 25 that its role may extend beyond regulation of lipid metabolism alone.

#### Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune disease, neurodegenerative disorders, 30 and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors 35 (Sanjo, H. et al. (1998) *J. Biol. Chem.* 273:29066-29071). DRAK1 and DRAK2 are STKs that share

homology with the death-associated protein kinases (DAP kinases), known to function in interferon- $\gamma$  induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with 5 apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving 10 the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 15 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the 20 CD95 apoptosis pathway (Inohara et al., *supra*).

#### Mitochondrial Protein Kinases

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity 25 with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, 30 regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) Adv. Enzyme Regul. 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of 35 the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes

in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) supra).

### KINASES WITH NON-PROTEIN SUBSTRATES

5

#### Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell.*

10 *Biol.* 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state ( $\text{PIP}_2$ ).  $\text{PIP}_2$  is then cleaved into inositol triphosphate ( $\text{IP}_3$ ) and diacylglycerol. These two products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate ( $\text{PIP}_2$ ) to PI (3,4,5) P<sub>3</sub> ( $\text{PIP}_3$ ).  $\text{PIP}_3$  then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) *Biochem. J.* 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, supra).

An example of lipid kinase phosphorylation activity is the phosphorylation of D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) *J. Biol. Chem.* 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein

coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including 5 platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., supra).

#### Purine Nucleotide Kinases

10 The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of 15 various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production 20 and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its 25 activity to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the 30 regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21<sup>ras</sup> known to be involved in control of cell proliferation and oncogenesis (Bos, J.L. (1989) Cancer Res. 49:4682-4689). High 35 ratios of GTP:GDP caused by suppression of GuK cause activation of p21<sup>ras</sup> and promote

oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buclovir (Miller, W.H. and R.L. Miller (1980) J. Biol. Chem. 255:7204-7207; Stenberg, K. et al. (1986) J. Biol. Chem. 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for reducing the necessary dosages of the drugs.

#### Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2. Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:11941-11945). Phosphorylation of deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) Pharmacol. Ther. 67:155-186).

The discovery of new human kinases and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of cancer, immune disorders, disorders affecting growth and development, atherosclerosis and other cardiovascular diseases, and lipid disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of human kinases.

#### **SUMMARY OF THE INVENTION**

The invention features purified polypeptides, human kinases, referred to collectively as "PKIN" and individually as "PKIN-1," "PKIN-2," "PKIN-3," "PKIN-4," "PKIN-5," "PKIN-6," "PKIN-7," "PKIN-8," "PKIN-9," "PKIN-10," and "PKIN-11." In one aspect, the invention provides an isolated polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-11.

- The invention further provides an isolated polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ 5 ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-11. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:12-22.
- 10 Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, 15 c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.
- 20 The invention also provides a method for producing a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of 25 SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.
- 30 Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence 35 selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino

acid sequence selected from the group consisting of SEQ ID NO:1-11.

The invention further provides an isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 5 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 10 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to 15 said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of 20 said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of a) a polynucleotide sequence selected from the group 25 consisting of SEQ ID NO:12-22, b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, c) a polynucleotide sequence complementary to a), d) a polynucleotide sequence complementary to b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) 30 detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

The invention further provides a composition comprising an effective amount of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino 35 acid sequence having at least 90% sequence identity to an amino acid sequence selected from the

group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence 5 selected from the group consisting of SEQ ID NO:1-11. The invention additionally provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) 10 an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The 15 method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such 20 treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence 25 selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising 30 an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds

to a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide comprising an amino acid sequence selected from the group consisting of a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:12-22, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a

polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence selected from the group consisting of i) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, ii) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22, iii) a polynucleotide sequence complementary to i), iv) a polynucleotide sequence complementary to ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

15

#### BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which  
5 will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so  
10 forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now  
15 described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

## DEFINITIONS

20 "PKIN" refers to the amino acid sequences of substantially purified PKIN obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of  
25 PKIN. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

An "allelic variant" is an alternative form of the gene encoding PKIN. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or  
30 many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding PKIN include those sequences with deletions,  
35 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as PKIN or a

polypeptide with at least one functional characteristic of PKIN. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding PKIN, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding

- 5 PKIN. The encoded protein may also be “altered,” and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent PKIN. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of PKIN is retained. For example,
- 10 negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

15 The terms “amino acid” and “amino acid sequence” refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where “amino acid sequence” is recited to refer to a sequence of a naturally occurring protein molecule, “amino acid sequence” and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

20 “Amplification” relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term “antagonist” refers to a molecule which inhibits or attenuates the biological activity of PKIN. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small  
25 molecules, or any other compound or composition which modulates the activity of PKIN either by directly interacting with PKIN or by acting on components of the biological pathway in which PKIN participates.

The term “antibody” refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')<sub>2</sub>, and Fv fragments, which are capable of binding an epitopic determinant.  
30 Antibodies that bind PKIN polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin,  
35 and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic PKIN, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 25 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding PKIN or fragments of PKIN may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated 35 DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied

Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and 5 assembled to produce the consensus sequence.

“Conservative amino acid substitutions” are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows 10 amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
15	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
20	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
25	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr
30		

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of 35 the side chain.

A “deletion” refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term “derivative” refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an 40 alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative

polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

5 A "fragment" is a unique portion of PKIN or the polynucleotide encoding PKIN which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10,  
10 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the  
15 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:12-22 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:12-22, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:12-22 is useful, for  
20 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:12-22 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:12-22 and the region of SEQ ID NO:12-22 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-11 is encoded by a fragment of SEQ ID NO:12-22. A fragment  
25 of SEQ ID NO:1-11 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-11. For example, a fragment of SEQ ID NO:1-11 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-11. The precise length of a fragment of SEQ ID NO:1-11 and the region of SEQ ID NO:1-11 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended  
30 purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between  
35 two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and  
5 therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in  
10 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS 8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.  
15

Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence  
20 analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST  
25 programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*  
*Reward for match: 1*  
30 *Penalty for mismatch: -2*  
*Open Gap: 5 and Extension Gap: 2 penalties*  
*Gap x drop-off: 50*  
*Expect: 10*  
*Word Size: 11*  
35 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases “percent identity” and “% identity,” as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and “diagonals saved”=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the “percent similarity” between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the “BLAST 2 Sequences” tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

*Matrix: BLOSUM62*

30           *Open Gap: 11 and Extension Gap: 1 penalties*

*Gap x drop-off: 50*

*Expect: 10*

*Word Size: 3*

*Filter: on*

35           Percent identity may be measured over the length of an entire defined polypeptide sequence,

for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment 5 length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

10 The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

15 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific 20 binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

25 Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point ( $T_m$ ) for the specific sequence at a defined ionic strength and pH. The  $T_m$  is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating  $T_m$  and 30 conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, 35 for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC

concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular 5 circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid 10 sequences by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g.,  $C_{0t}$  or  $R_{0t}$  analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

15 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect 20 cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of PKIN which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of PKIN which is useful in any of the antibody production methods disclosed herein or known in the 25 art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

30 The term "modulate" refers to a change in the activity of PKIN. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of PKIN.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, 35 polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the

antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition.

10 PNAS preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

15 "Post-translational modification" of an PKIN may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of PKIN.

"Probe" refers to nucleic acid sequences encoding PKIN, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

20 "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

25 Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

30 Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2<sup>nd</sup> ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs 35 can be derived from a known sequence, for example, by using computer programs intended for that

purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 5 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from 10 megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a “mispriming library,” in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection 15 programs may also be obtained from their respective sources and modified to meet the user’s specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both 20 unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

25 A “recombinant nucleic acid” is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have 30 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

35 Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is

expressed, inducing a protective immunological response in the mammal.

A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, 5 translation, or RNA stability.

“Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

10 An “RNA equivalent,” in reference to a DNA sequence, is composed of the same linear sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

15 The term “sample” is used in its broadest sense. A sample suspected of containing PKIN, nucleic acids encoding PKIN, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

20 The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

25 The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

30 A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

“Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

35 A “transcript image” refers to the collective pattern of gene expression by a particular cell

type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to

another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The 5 presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-10 1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

## 15 THE INVENTION

The invention is based on the discovery of new human kinases (PKIN), the polynucleotides encoding PKIN, and the use of these compositions for the diagnosis, treatment, or prevention of cancer, immune disorders, disorders affecting growth and development, atherosclerosis and other cardiovascular diseases, and lipid disorders.

20 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is 25 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the 30 polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

35 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and

2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the  
5 MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these  
10 properties establish that the claimed polypeptides are human kinases. For example, SEQ ID NO:5 is 81% identical, from residue M1 to residue Q783, to rat salt-inducible protein kinase (GenBank ID g5672676) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is negligible, indicating the probability of obtaining the observed  
15 polypeptide sequence alignment by chance. SEQ ID NO:5 also contains protein kinase domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS,  
MOTIFS, and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:5 is a protein kinase. SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:6, SEQ ID  
20 NO:7, SEQ ID NO:8, SEQ ID NO:9, SEQ ID NO:10, and SEQ ID NO:11 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-11 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence  
25 identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:12-22 or that distinguish between SEQ ID NO:12-22 and  
30 related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5')  
35 and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 7286907H1 is the identification number of an Incyte cDNA sequence, and BRAIFER06 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled 5 cDNA libraries (e.g., 70067849V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GNN.g6560812\_002.edit is the identification number of a Genscan-predicted coding sequence, with g6560812 being the GenBank 10 identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted 15 exons brought together by an “exon stitching” algorithm. For example, FL6756529\_00001 represents a “stitched” sequence in which 6756529 is the identification number of the cluster of sequences to which the algorithm was applied, and 00001 is the number of the prediction generated by the algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an “exon-stretching” 20 algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors 25 which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses PKIN variants. A preferred PKIN variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the PKIN amino acid sequence, and which contains at least one functional or structural characteristic of PKIN.

30 The invention also encompasses polynucleotides which encode PKIN. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22, which encodes PKIN. The polynucleotide sequences of SEQ ID NO:12-22, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone 35 is composed of ribose instead of deoxyribose.

- The invention also encompasses a variant of a polynucleotide sequence encoding PKIN. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding PKIN. A particular aspect of the invention encompasses a variant of a
- 5 polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:12-22 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:12-22. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of PKIN.
- 10 It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding PKIN, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These
- 15 combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring PKIN, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode PKIN and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring PKIN under appropriately selected

20 conditions of stringency, it may be advantageous to produce nucleotide sequences encoding PKIN or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide

25 sequence encoding PKIN and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences which encode PKIN and PKIN derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the

30 synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding PKIN or any fragment thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID

35 NO:12-22 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and

S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of 5 the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is 10 automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences 15 are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

The nucleic acid sequences encoding PKIN may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, 20 such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) PCR Methods Applic. 2:318-322.) Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising 25 a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119.) In this method, multiple restriction enzyme 30 digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed 35 using commercially available software, such as OLIGO 4.06 primer analysis software (National

Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been  
5 size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze  
10 the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire  
15 process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof  
which encode PKIN may be cloned in recombinant DNA molecules that direct expression of PKIN,  
20 or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express PKIN.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter PKIN-encoding sequences for a variety of purposes including, but  
25 not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

30 The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of PKIN, such as its biological or enzymatic activity or its ability to  
35 bind to other molecules or compounds. DNA shuffling is a process by which a library of gene

variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" 5 breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable 10 manner.

In another embodiment, sequences encoding PKIN may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.) Alternatively, PKIN itself or a fragment thereof may be synthesized using chemical methods. For 15 example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) Science 269:202-204.) Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of PKIN, or any part thereof, may be altered during direct synthesis and/or 20 combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by 25 sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active PKIN, the nucleotide sequences encoding PKIN or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and 30 inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding PKIN. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences encoding PKIN. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding PKIN and its initiation codon and upstream regulatory sequences are inserted 35 into the appropriate expression vector, no additional transcriptional or translational control signals

may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of  
5 enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding PKIN and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in  
10 vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding PKIN. These include, but are not limited to, microorganisms such as bacteria transformed  
15 with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, supra; Ausubel, supra; Van Heeke, G. and S.M. Schuster  
20 (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses,  
25 adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al.  
30 (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)  
The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotide sequences encoding PKIN. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding PKIN can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1  
35 plasmid (Life Technologies). Ligation of sequences encoding PKIN into the vector's multiple

cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of PKIN are needed, e.g. for the production of antibodies, vectors which direct high level expression of PKIN may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of PKIN. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel, 1995, supra; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; and Scorer, C.A. et al. (1994) Bio/Technology 12:181-184.)

Plant systems may also be used for expression of PKIN. Transcription of sequences encoding PKIN may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding PKIN may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses PKIN in host cells. (See, e.g., Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 35: 100-104.)

15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of PKIN in cell lines is preferred. For example, sequences encoding PKIN can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*<sup>-</sup> and *apr*<sup>-</sup> cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech),  $\beta$  glucuronidase and its substrate  $\beta$ -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding PKIN is inserted within a marker gene sequence, transformed cells containing sequences encoding PKIN can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding PKIN under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding PKIN and that express PKIN may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR

amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of PKIN using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques 5 include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on PKIN is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, 10 Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled 15 hybridization or PCR probes for detecting sequences related to polynucleotides encoding PKIN include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding PKIN, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase 20 such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

25 Host cells transformed with nucleotide sequences encoding PKIN may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode PKIN may be designed to contain signal sequences which 30 direct secretion of PKIN through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or 35 "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity.

Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

- 5 In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding PKIN may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric PKIN protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of PKIN activity. Heterologous protein and  
10 peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and  
15 metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the PKIN encoding sequence and the heterologous protein sequence, so that PKIN may be cleaved away from the heterologous moiety following purification.  
20 Methods for fusion protein expression and purification are discussed in Ausubel (1995, supra, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled PKIN may be achieved in vitro using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These  
25 systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, <sup>35</sup>S-methionine.

PKIN of the present invention or fragments thereof may be used to screen for compounds that specifically bind to PKIN. At least one and up to a plurality of test compounds may be screened for  
30 specific binding to PKIN. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of PKIN, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):  
35 Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which PKIN

binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express PKIN, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, Drosophila, or 5 E. coli. Cells expressing PKIN or cell membrane fractions which contain PKIN are then contacted with a test compound and binding, stimulation, or inhibition of activity of either PKIN or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, 10 the assay may comprise the steps of combining at least one test compound with PKIN, either in solution or affixed to a solid support, and detecting the binding of PKIN to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical 15 libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

PKIN of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of PKIN. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for PKIN activity, wherein PKIN is combined with at least one test compound, and the activity of PKIN in the 20 presence of a test compound is compared with the activity of PKIN in the absence of the test compound. A change in the activity of PKIN in the presence of the test compound is indicative of a compound that modulates the activity of PKIN. Alternatively, a test compound is combined with an in vitro or cell-free system comprising PKIN under conditions suitable for PKIN activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of PKIN 25 may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding PKIN or their mammalian homologs may be “knocked out” in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal 30 models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region

of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell 5 blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding PKIN may also be manipulated in vitro in ES cells derived from 10 human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding PKIN can also be used to create "knockin" humanized animals 15 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding PKIN is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. 20 Alternatively, a mammal inbred to overexpress PKIN, e.g., by secreting PKIN in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

## THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists 25 between regions of PKIN and human kinases. In addition, the expression of PKIN is closely associated with cancer, including breast and prostate cancer. Therefore, PKIN appears to play a role in cancer, immune disorders, disorders affecting growth and development, atherosclerosis and other cardiovascular diseases, and lipid disorders. In the treatment of disorders associated with increased PKIN expression or activity, it is desirable to decrease the expression or activity of PKIN. In the treatment of disorders associated with decreased PKIN expression or activity, it is desirable to 30 increase the expression or activity of PKIN.

Therefore, in one embodiment, PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in 35 particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall

bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

5 allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's

10 syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome,

15 complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma,

20 leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy,

25 gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and

30 sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease,

35 degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve,

mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, 5 pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, 10 pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, 15 hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, 20 lipodystrophy, lipomatosis, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous 25 xanthomatosis, sitosterolemia, hypcholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity.

In another embodiment, a vector capable of expressing PKIN or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those described above.

30 In a further embodiment, a composition comprising a substantially purified PKIN in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of PKIN including, but not limited to, those provided above.

35 In still another embodiment, an agonist which modulates the activity of PKIN may be administered to a subject to treat or prevent a disorder associated with decreased expression or

activity of PKIN including, but not limited to, those listed above.

In a further embodiment, an antagonist of PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN. Examples of such disorders include, but are not limited to, those cancer, immune disorders, disorders affecting growth and development, atherosclerosis and other cardiovascular diseases, and lipid disorders, described above. In one aspect, an antibody which specifically binds PKIN may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express PKIN.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding PKIN may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of PKIN including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of PKIN may be produced using methods which are generally known in the art. In particular, purified PKIN may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind PKIN. Antibodies to PKIN may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with PKIN or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to PKIN have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or

fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of PKIN amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to PKIN may be prepared using any technique which provides for the 5 production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030; and Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120.)

10 In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; and Takeda, S. et al. (1985) *Nature* 314:452-454.) Alternatively, techniques described for the production of single 15 chain antibodies may be adapted, using methods known in the art, to produce PKIN-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137.)

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte 20 population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299.)

Antibody fragments which contain specific binding sites for PKIN may also be generated. 25 For example, such fragments include, but are not limited to, F(ab')<sub>2</sub> fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')<sub>2</sub> fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) *Science* 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired 30 specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between PKIN and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering PKIN epitopes is generally used, but a competitive binding assay may 35 also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for PKIN. Affinity is expressed as an association constant,  $K_a$ , which is defined as the molar concentration of PKIN-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.

5 The  $K_a$  determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple PKIN epitopes, represents the average affinity, or avidity, of the antibodies for PKIN. The  $K_a$  determined for a preparation of monoclonal antibodies, which are monospecific for a particular PKIN epitope, represents a true measure of affinity. High-affinity antibody preparations with  $K_a$  ranging from about  $10^9$  to  $10^{12}$  L/mole are preferred for use in immunoassays in which the

10 PKIN-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with  $K_a$  ranging from about  $10^6$  to  $10^7$  L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of PKIN, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons,

15 New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of PKIN-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g., Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding PKIN, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding PKIN. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding PKIN. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totowa NJ.)

30 In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Cli. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 35 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral

vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) Blood 76:271; Ausubel, *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) Br. Med. Bull. 51(1):217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87(11):1308-1315; and Morris, M.C. et al. (1997) Nucleic Acids Res. 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding PKIN may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in PKIN expression or regulation causes disease, the expression of PKIN from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in PKIN are treated by constructing mammalian expression vectors encoding PKIN and introducing these vectors by mechanical means into PKIN-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of PKIN include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA),

PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). PKIN may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or  $\beta$ -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding PKIN from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to PKIN expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding PKIN under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4 $^{+}$  T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in

the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

5 In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding PKIN to cells which have one or more genetic abnormalities with respect to the expression of PKIN. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas  
10 (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999) Annu. Rev. Nutr. 19:511-544 and Verma, I.M. and N. Somia (1997) Nature 18:389:239-242, both incorporated by reference herein.

15 In another alternative, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding PKIN to target cells which have one or more genetic abnormalities with respect to the expression of PKIN. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing PKIN to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant  
20 HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) J. Virol. 73:519-532 and Xu, H. et al. (1994) Dev. Biol. 163:152-161, hereby incorporated by reference. The manipulation of cloned  
25 herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

30 In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding PKIN to target cells. The biology of the prototypic alphavirus,

Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA,

5 resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for PKIN into the alphavirus genome in place of the capsid-coding region results in the production of a large number of PKIN-coding RNAs and the synthesis of high levels of PKIN in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a

10 persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of PKIN into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of

15 manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfactions, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly,

20 inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-

25 177.) A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example,

30 engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding PKIN.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides,

35 corresponding to the region of the target gene containing the cleavage site, may be evaluated for

secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared  
5 by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding PKIN. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs  
10 that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase  
15 linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

20 An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding PKIN. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular  
25 chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased PKIN expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding PKIN may be therapeutically useful, and in the treatment of disorders associated with decreased PKIN expression or activity, a compound which specifically promotes expression of the polynucleotide encoding PKIN may be therapeutically useful.  
30

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in  
35 altering polynucleotide expression; selection from an existing, commercially-available or proprietary

library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding PKIN is exposed to at least one test compound thus obtained. The sample  
5 may comprise, for example, an intact or permeabilized cell, or an in vitro cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding PKIN are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding PKIN. The amount of hybridization may be quantified, thus forming  
10 the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a Schizosaccharomyces pombe gene expression system  
15 (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a  
20 combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient.  
25 Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and  
30 monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of  
35 Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may

consist of PKIN, antibodies to PKIN, and mimetics, agonists, antagonists, or inhibitors of PKIN.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, 5 enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides 10 and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active 15 ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of 20 macromolecules comprising PKIN or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, PKIN or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell 25 culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example PKIN 30 or fragments thereof, antibodies of PKIN, and agonists, antagonists or inhibitors of PKIN, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) or LD<sub>50</sub> (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the 35 therapeutic index, which can be expressed as the LD<sub>50</sub>/ED<sub>50</sub> ratio. Compositions which exhibit large

therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the 5 patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the 10 subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 µg to 100,000 µg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and 15 methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

## DIAGNOSTICS

20 In another embodiment, antibodies which specifically bind PKIN may be used for the diagnosis of disorders characterized by expression of PKIN, or in assays to monitor patients being treated with PKIN or agonists, antagonists, or inhibitors of PKIN. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for PKIN include methods which utilize the antibody and a label to detect PKIN in human body fluids 25 or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring PKIN, including ELISAs, RIAs, and FACS, are known 30 in the art and provide a basis for diagnosing altered or abnormal levels of PKIN expression. Normal or standard values for PKIN expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to PKIN under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of PKIN expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation 35 between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding PKIN may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of PKIN may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of PKIN, and to monitor regulation of PKIN levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding PKIN or closely related molecules may be used to identify nucleic acid sequences which encode PKIN. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding PKIN, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the PKIN encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:12-22 or from genomic sequences including promoters, enhancers, and introns of the PKIN gene.

Means for producing specific hybridization probes for DNAs encoding PKIN include the cloning of polynucleotide sequences encoding PKIN or PKIN derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as <sup>32</sup>P or <sup>35</sup>S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding PKIN may be used for the diagnosis of disorders associated with expression of PKIN. Examples of such disorders include, but are not limited to, a cancer, such as adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, leukemias such as multiple myeloma and lymphomas such as Hodgkin's disease; an immune disorder, such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's

disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a growth and developmental disorder, such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Syndenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a cardiovascular disease, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, 25 thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, 30 infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease,

emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia,

5 diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; and a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine

10 deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM<sub>2</sub> gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid

15 adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity. The polynucleotide sequences encoding

20 PKIN may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered PKIN expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding PKIN may be useful in assays that

25 detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding PKIN may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to

30 a control sample then the presence of altered levels of nucleotide sequences encoding PKIN in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of PKIN,

35 a normal or standard profile for expression is established. This may be accomplished by combining

body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding PKIN, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide 5 is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the 10 patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the 15 development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding PKIN 20 may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced in vitro. Oligomers will preferably contain a fragment of a polynucleotide encoding PKIN, or a fragment of a polynucleotide complementary to the polynucleotide encoding PKIN, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or 25 quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded 30 conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding PKIN are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and 35 these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the

oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of PKIN include radiolabeling or biotinylation of nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, PKIN, fragments of PKIN, or antibodies specific for PKIN may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number

5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present 5 invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo, as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

10 Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and 15 toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide 20 measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, 25 knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

30 In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with 35 levels in an untreated biological sample. Differences in the transcript levels between the two samples

are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a  
5 proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are  
10 separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson,  
supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The  
15 optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the  
spots are partially sequenced using, for example, standard methods employing chemical or enzymatic  
cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by  
20 comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the  
polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for PKIN to quantify the levels of PKIN expression. In one embodiment, the antibodies are used as elements on a microarray,  
25 and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-  
111; Mendoza, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol-  
or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each  
30 array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson,  
N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be  
35 useful in the analysis of compounds which do not significantly affect the transcript image, but which

alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

In another embodiment of the invention, nucleic acid sequences encoding PKIN may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 35: Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet.

7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

5       Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding PKIN on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the 10 region of DNA associated with that disorder and thus may further positional cloning efforts.

*In situ* hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is 15 valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of 20 the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

      In another embodiment of the invention, PKIN, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a 25 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between PKIN and the agent being tested may be measured.

      Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are 30 synthesized on a solid substrate. The test compounds are reacted with PKIN, or fragments thereof, and washed. Bound PKIN is then detected by methods well known in the art. Purified PKIN can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

      In another embodiment, one may use competitive drug screening assays in which neutralizing 35 antibodies capable of binding PKIN specifically compete with a test compound for binding PKIN. In

this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with PKIN.

In additional embodiments, the nucleotide sequences which encode PKIN may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure 10 in any way whatsoever.

The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Serial No. 60/183,682, U.S. Serial No. 60/186,559, U.S. Serial No. 60/188,606, U.S. Serial No. 60/189,998, and U.S. Serial No. 60/193,851, are expressly incorporated by reference herein.

15

## EXAMPLES

### I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized 20 and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

25 Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA 30 purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 35 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic

oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs 5 were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid (Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 $\alpha$ , DH10B, or ElectroMAX 10 DH10B from Life Technologies.

## II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an 15 AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a 20 high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

## 25 III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the 30 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the 35 ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI

protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, *supra*, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

- 5       The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and
- 10     BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.
- 15     Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive
- 20     the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length
- 25     polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

- 30     Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the

strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ 5 ID NO:12-22. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

#### **IV. Identification and Editing of Coding Sequences from Genomic DNA**

Putative human kinases were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (See Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94, and Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for 10 Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode human kinases, the encoded polypeptides were analyzed by querying against 15 PFAM models for human kinases. Potential human kinases were also identified by homology to Incyte cDNA sequences that had been annotated as human kinases. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public 20 databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of 25 the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

#### **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**

##### **"Stitched" Sequences**

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm 35 based on graph theory and dynamic programming to integrate cDNA and genomic information,

generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then “stitched” together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

15     “Stretched” Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore “stretched” or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

25     **VI. Chromosomal Mapping of PKIN Encoding Polynucleotides**

The sequences which were used to assemble SEQ ID NO:12-22 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:12-22 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences

had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, or human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

In this manner, SEQ ID NO:20 was mapped to chromosome 1 within the interval from 75.30 to 81.60 centiMorgans.

## 15 VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, supra, ch. 7; Ausubel (1995) supra, ch. 4 and 16.)

20 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

25

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate

the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the  
5 other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotide sequences encoding PKIN are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA  
10 sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed;  
15 or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and  
20 disease-specific expression of cDNA encoding PKIN. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

### VIII. Extension of PKIN Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One  
25 primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would  
30 result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction  
35 mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg<sup>2+</sup>, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>,

and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C. In the 5 alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 57 °C, 1 min; Step 4: 68 °C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68 °C, 5 min; Step 7: storage at 4 °C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE 10 and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the 15 sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) 20 agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent E. coli cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37 °C in 25 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 72 °C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72 °C, 5 min; Step 7: storage at 4 °C. DNA was 30 quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

35 In like manner, full length polynucleotide sequences are verified using the above procedure or

are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

#### IX. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:12-22 are employed to screen cDNAs, 5 genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250  $\mu$ Ci of [ $\gamma$ - $^{32}$ P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase 10 (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech). An aliquot containing  $10^7$  counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

15 The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and 20 compared.

#### X. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), 25 mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link 30 elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may 35 comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The

array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

#### Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)<sup>+</sup> RNA is purified using the oligo-(dT) cellulose method. Each poly(A)<sup>+</sup> RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ $\mu$ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ $\mu$ l RNase inhibitor, 500  $\mu$ M dATP, 500  $\mu$ M dGTP, 500  $\mu$ M dTTP, 40  $\mu$ M dCTP, 40  $\mu$ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)<sup>+</sup> RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)<sup>+</sup> RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc. (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14  $\mu$ l 5X SSC/0.2% SDS.

#### Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5  $\mu$ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia Biotech).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

- Array elements are applied to the coated glass substrate using a procedure described in US 5 Patent No. 5,807,522 , incorporated herein by reference. 1  $\mu$ l of the array element DNA, at an average concentration of 100 ng/ $\mu$ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. 10 Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

#### Hybridization

Hybridization reactions contain 9  $\mu$ l of sample mixture consisting of 0.2  $\mu$ g each of Cy3 and 15 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm<sup>2</sup> coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140  $\mu$ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is 20 incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

#### Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an 25 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a 30 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The

emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

- The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

- A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

## XI. Complementary Polynucleotides

Sequences complementary to the PKIN-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring PKIN. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of PKIN. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the PKIN-encoding transcript.

## XII. Expression of PKIN

Expression and purification of PKIN is achieved using bacterial or virus-based expression systems. For expression of PKIN in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA

transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3).

Antibiotic resistant bacteria express PKIN upon induction with isopropyl beta-D-

- 5 thiogalactopyranoside (IPTG). Expression of PKIN in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant Autographica californica nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding PKIN by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong 10 polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

- 15 In most expression systems, PKIN is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham 20 Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from PKIN at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, 25 supra, ch. 10 and 16). Purified PKIN obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

### XIII. Functional Assays

- PKIN function is assessed by expressing the sequences encoding PKIN at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression 30 vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences encoding a 35 marker protein are co-transfected. Expression of a marker protein provides a means to distinguish

transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate 5 the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; 10 alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of PKIN on gene expression can be assessed using highly purified populations 15 of cells transfected with sequences encoding PKIN and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. 20 Expression of mRNA encoding PKIN and other genes of interest can be analyzed by northern analysis or microarray techniques.

#### XIV. Production of PKIN Specific Antibodies

PKIN substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to 25 immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the PKIN amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well 30 described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the 35 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for

antipeptide and anti-PKIN activity by, for example, binding the peptide or PKIN to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

#### XV. Purification of Naturally Occurring PKIN Using Specific Antibodies

- 5       Naturally occurring or recombinant PKIN is substantially purified by immunoaffinity chromatography using antibodies specific for PKIN. An immunoaffinity column is constructed by covalently coupling anti-PKIN antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.
- 10      Media containing PKIN are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of PKIN (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/PKIN binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and PKIN is collected.

#### 15 XVI. Identification of Molecules Which Interact with PKIN

- PKIN, or biologically active fragments thereof, are labeled with <sup>125</sup>I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled PKIN, washed, and any wells with labeled PKIN complex are assayed. Data obtained using different concentrations 20 of PKIN are used to calculate values for the number, affinity, and association of PKIN with the candidate molecules.

Alternatively, molecules interacting with PKIN are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989) Nature 340:245-246, or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

25      PKIN may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

#### XVII. Demonstration of PKIN Activity

- 30      Generally, protein kinase activity is measured by quantifying the phosphorylation of a protein substrate by PKIN in the presence of gamma-labeled <sup>32</sup>P-ATP. PKIN is incubated with the protein substrate, <sup>32</sup>P-ATP, and an appropriate kinase buffer. The <sup>32</sup>P incorporated into the substrate is separated from free <sup>32</sup>P-ATP by electrophoresis and the incorporated <sup>32</sup>P is counted using a radioisotope counter. The amount of incorporated <sup>32</sup>P is proportional to the activity of PKIN. A 35 determination of the specific amino acid residue phosphorylated is made by phosphoamino acid

analysis of the hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma  $^{32}\text{P}$ -ATP. Following the reaction, free avidin in solution is added for binding to the biotinylated  $^{32}\text{P}$ -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma  $^{32}\text{P}$ -ATP. The reservoir of the centrifuged unit containing the  $^{32}\text{P}$ -peptide product as retentate is then counted in a scintillation counter. This procedure allows assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation, Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes are as follows: Histone H1 (Sigma) and p34<sup>cdk2</sup>kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

In another alternative, protein kinase activity of PKIN is demonstrated in vitro in an assay containing PKIN, 50 $\mu\text{l}$  of kinase buffer, 1 $\mu\text{g}$  substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10  $\mu\text{g}$  ATP, and 0.5 $\mu\text{Ci}$  [ $\gamma$ - $^{33}\text{P}$ ]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [ $\gamma$ - $^{33}\text{P}$ ]ATP is removed by washing and the incorporated radioactivity is measured using a radioactivity scintillation counter. Alternatively, the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and visualized on a 12% SDS polyacrylamide gel by autoradiography. Incorporated radioactivity is corrected for reactions carried out in the absence of PKIN or in the presence of the inactive kinase, K38A.

In yet another alternative, adenylate kinase or guanylate kinase activity may be measured by the incorporation of  $^{32}\text{P}$  from gamma-labeled  $^{32}\text{P}$ -ATP into ADP or GDP using a gamma radioisotope counter. The enzyme, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and  $^{32}\text{P}$ -labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is cut out and counted. The radioactivity recovered is proportional to the enzyme activity.

In yet another alternative, other assays for PKIN include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin.

Inhibitors of PKIN activity, such as small organic molecules, proteins or peptides, may be identified by such assays.

#### XVIII. Enhancement/Inhibition of Protein Kinase Activity

Agonists or antagonists of PKIN activation or inhibition may be tested using assays described  
5 in section XVII. Agonists cause an increase in PKIN activity and antagonists cause a decrease in  
PKIN activity.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the  
10 invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polymucleotide SEQ ID NO:	Incyte Polymucleotide ID
4912053	1	4912053CD1	12	4912053CB1
870615	2	870615CD1	13	870615CB1
2762870	3	2762870CD1	14	2762870CB1
5507012	4	5507012CD1	15	5507012CB1
6756529	5	6756529CD1	16	6756529CB1
387382	6	387382CD1	17	387382CB1
5850786	7	5850786CD1	18	5850786CB1
7182343	8	7182343CD1	19	7182343CB1
7264345	9	7264345CD1	20	7264345CB1
7472328	10	7472328CD1	21	7472328CB1
1471412	11	1471412CD1	22	1471412CB1

**Table 2**

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
1	4912053CD1	g190660	1.0e-06	Protein-serine kinase [Homo sapiens]
2	870615CD1	g340026	1.00E-19	Tyrosine kinase [Homo sapiens] Takahashi, M. and Cooper, G.M. (1987) Mol. Cell. Biol. 7:1378-1385.
3	2762870CD1	g4376414	5.9e-31	Serine/Threonine Protein Kinase [Chlamydophila pneumoniae]
4	5507012CD1	g4995303	1.6e-134	putative mitogen-activated protein kinase kinase kinase [Homo sapiens]
		g8272557	0.0	protein kinase WNK1 [Rattus norvegicus] Xu, B. et al. (2000) J. Biol. Chem. 275:16795-16801.
5	6756529CD1	g5672676	0.0	salt-inducible protein kinase [Rattus norvegicus]
6	387382CD1	g4519166	6.5e-209	GSG2 [Mus musculus] Tanaka, H. et al. (1994) FEBS Lett. 355: 4-10.
7	5850786CD1	g4101587	2.1e-133	serine/threonine kinase [Homo sapiens]
		g7649810	0	protein kinase PAK5 [Homo sapiens]
8	7182343CD1	g182571	3.6e-34	Fibroblast growth factor receptor 4 [Homo sapiens]
9	7264345CD1	g4753775	3.7e-300	protein kinase [Homo sapiens] Rosok, O. et al. (1999) Genomics 61: 44-54
10	7472328CD1	g4827159	3.9e-110	kinase [Rattus sp.] Matsuda, K.M. et al. (1999) Biochem. Biophys. Res. Commun. 258: 260-264
11	1471412CD1	g2393890	1.8e-26	protein kinase homolog [Fowlpox virus]

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	4912053CD1	792	S9 T19 S66 T104 T260 T303 T471 T521 S523 T542 T585 T623 S540 S30 T172 S346 T418 T575 T600 T628	N46 N493 N589 N631 N779	PROTEIN SCY1: PD043026: T193-H469 Eukaryotic protein kinase domain: V20-E186 Tyrosine kinase catalytic domain signature: PRO00109: L27-L45, S106-V128	BLAST-PRODOM HMMER-PFAM BLIMPS-PRINTS
2	870615CD1	224	S19 T50 S214 T128 S131 S191	N156	Domain found in NIK1-like kinase: PF00780:E169-V177	BLIMPS-PFAM
3	2762870CD1	374	T120 S132 T134 T147 S210 T227 S234 T270 S311 S323	N141	Signal peptide: M1-A32 Signal cleavage: M1-S34 PROTEIN KINASE YFMG SERINE/THREONINE SERINE/THREONINE PROTEIN KINASE [Chlamydophila pneumoniae]: PD017761: W179- D368	HMMER SPSCAN BLAST-PRODOM
4	5507012CD1	565	T60 T258 S378 T436 S29 S34 T73 T91 T160 S174 S189 T243 T258 S469 T48 S231 S260 T290 T308 S363 T373 Y468	N27 N89	Serine/Threonine protein kinases active- site signature: I345-I357 Eukaryotic protein kinase domain: L221-F479 Protein_kinases_tyrosine signature and profile: L324-F379 Tyrosine kinase catalytic domain signature: PRO00109A:T301-K314; PRO00109B: H339-I357; PRO00109C: V403-C425; PRO00109D: A448-I470 Protein Kinase Domain: DM00004   S49611   39- 259:I227-V447	MOTIFS HMMER-PFAM PROFILESCAN BLIMPS-PRINTS BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	6756529CD1	783	T268 T473 S575 T594 T61 S66 S122 S135 S370 S438 T473 S498 T503 S567 S649 T231 T468 S478 T593 Y114	N328	Protein kinases ATP-binding region signature: L33-K56 Serine/Threonine protein kinases active-site signature: I145-L157 Eukaryotic protein kinase domain: Y27-M278, Y304-E343, V397-S424 Protein_kinases_tyrosine signature and PROFILESCAN profile D97-G177	MOTIFS
6	387382CD1	798	T295 S458 S661 S53 S58 T181 S220 T253 S299 S458 T501 T528 S586 S684 S691 S705 S786 T97 T112 S143 S240 T289 S336 S337 S353 S421 S453 S468 T481 S582 T624 S642 T749 T757 Y739	N45 N427 N522 N551 N581 N783	Tyrosine kinase catalytic domain signature: BLIMPS_PRINTS PR00109A: T103-T116; PR00109B: Y139-L157; PR00109C: I205-P227 Protein Kinase Domain: DM00004   P27448   58 - BLAST-DOMO 297:R31-I269 Protein Kinase Atp: I490-K511	MOTIFS

Table 3 (cont.)

SEQ NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
7	5850786CD1	632	S104 S15 T23 S133 T137 S201 S226 S253 S260 S345 S418 S441 T603 S99 T143 T281 T467 T471 S599 Y152 Y182	N91 N274	Protein Kinase Atp: I455-K478 Eukaryotic protein kinase domain: L449-V582, L586-L613 P21-Rho-binding domain: I11-T69 Tyrosine kinase catalytic domain: PRO0109B:Y558-L576	MOTIFS HMMER-PFAM
8	7182343CD1	350	T51 T38 T102 S140 S159 T268 S336 S13 S53 S62 T66 S70 T79 S193 Y65 Y182	N77 N101 N139	Transmembrane domain: V226-C245 Immunoglobulin domain: G11-V69, G107-G186 Receptor tyrosine kinase: BL00240B L16-R39	HMMER-PFAM BLIMPS-BLOCKS
8	7182343CD1	350	T51 T38 T102 S140 S159 T268 S336 S13 S53 S62 T66 S70 T79 S193 Y65 Y182	N77 N101 N139	BASIC FIBROBLAST GROWTH FACTOR RECEPTOR DM01287 S42803 1-713 :M1-R264 DM01287 S19947 1-822 :M1-R264 DM01287 A60350 1-800 :M1-Q303 DM01287 JC1450 1-650 :M1-P203	BLAST-DOMO
9	7264345CD1	571	S8 S345 T54 S301 T308 S345 S346 S367 S381 S477 S4 S42 S93 T151 T178 S211 T330 S373 T396 S403 S488 Y191 Y259	N31	Protein Kinase Atp: I64-K87 Protein Kinase Tyr: I172-I184 Eukaryotic protein kinase domain: E62-I305 Protein kinases signatures and profile: V152-D206	MOTIFS
9	7264345CD1	571	S8 S345 T54 S301 T308 S345 S346 S367 S381 S477 S4 S42 S93 T151 T178 S211 T330 S373 T396 S403 S488 Y191 Y259	N31	Tyrosine kinase catalytic domain: PRO0109A:T130-D143, PR00109B:Y166-I184 PRO0109C:G218-L228, PR00109D:A237-Y259, PRO0109E:C280-F302	BLIMPS-PRINTS
9	7264345CD1	571	S8 S345 T54 S301 T308 S345 S346 S367 S381 S477 S4 S42 S93 T151 T178 S211 T330 S373 T396 S403 S488 Y191 Y259	N31	Protein KINASE DOMAIN DM00004 I48737 341-596 :E62-F302 DM00004 P53667 340-596 :E62-F302 DM00004 P53666 332-593 :E62-F273 DM00004 S49313 23-266 :D58-F302	BLAST-DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	7472328CD1	323	T175 S180 T140 T27 T95 T159 S197 S203	T54 T90 T140	Eukaryotic protein kinase domain: P188-L280	HMMER-PFAM BLAST-DOMO
11	1471412CD1	471	S106 S467 T290 T363 S378 S458 S55 S80 S161 T175 T253 T368 S417 S454 Y23	N81	Eukaryotic protein kinase domain: Y218-K464	HMMER-PFAM Tyrosine kinase catalytic domain signature BLIMPS-PRINTS PRO0109: M283-D296; S393-C415; C437-V459 PROTEIN KINASE DOMAIN DM00004   JQ1743   73 - 262: N261-V459

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
12	4912053CB1	4335	1-592, 4026-4335, 3351-3496, 1123-2842	70067849V1 7286907H1 (BRAIFER06) 4912053H1 (THYMDIT01)	1715 784 230	2313 1323 514
				2659638F6 (LUNGUT09) 2545191T6 (UTRSNOT11) 1406536F6 (LATRTUT02)	3361 3666 2478	3875 4131 2996
				2545191F6 (UTRSNOT11) 70066706V1 6526407H1 (CONFNOT07)	3064 1854 1315	3477 2419 1855
				6916054H1 (PLACFER06) 2874082T6 (THYRNOT10)	256 2717	845 3334
				GNN.q6560812_002.edit 2550210F6 (LUNGUT06) 2874082F6 (THYRNOT10)	1 3837 1143	546 4335 1756
				70063710V1 6583325H1 (LUNGFFC01) 70882046V1	2321 1 620	2841 597 1201
				70885573V1 2172659F6 (ENDCNOT03)	558 1576	1169 1828
				70485671V1 70483457V1	1237 1115	1729 1655
				7151622H1 (BONEUNR01) 6308250H1 (NERDTDN03)	646 410	1168 972
				6545148H1 (LNODNON02) 4298443H1 (SCOMDIT01)	1201 1888	1762 2143
				4548711T1 (HELAUNT01)	1474	2117
				6486245H1 (MIXDUNB01) 6780258H1 (OVARDIR01)	993 1	1599 572
				2058456R6 (OVARNOT03)	1772	2126

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
15	5507012CB1	2310	1-158, 1433-1571, 671-802	70294304D1 7312202H1 6771174J1 (BRAUNOR01) 6808373H1 (SKIRNOR01) 70425501D1 7088726H1 (BRAUTDR03)	1809 1 770 359 1883 1511	2231 380 1529 865 2310 2094
16	6756529CB1	4699	3346-3845, 1-2800	1926322R6 (BRSTNOT02) 6932134H1 (SINTTMR02) 1963560R6 (BRSTNOT04) 1861375H1 (PROSNOT19) FL6756529_00001 1858047H1 (PROSNOT18) 1995450T6 (BRSTTUT03)	4366 1 3544 3510 664 3218 3966	4699 701 4057 3778 3008 3497 4672
17	387382CB1	2649	1-372, 2629-2649, 590-1733	1657259H1 (URETTUT01) 5617980H1 (THYMNOR02) 3330722H1 (HEAONOT04) 1699226F6 (BLADTUT05) 1746121H1 (STOMTUT02) 387382T6 (THYMNOT02)	3063 2724 3272 4259 2959 2381	3243 3019 3551 4679 3239 2649
18	5850786CB1	2526	1903-1997, 562-1537	GNN.96006112_000006_002 387382R6 (THYMNNOT02) 5262448T6 (CONDTUT02) 5262448F6 (CONDTUT02)	17 1937 1824 1	2448 2649 2519 493
				7159938H1 (HNT2TXC01) 2047113H1 (THP1T7T01) 4849452F9 (SPLNTUT02)	1562 1 303	1251 2526 421 1024
				GNN.96006112_000006_002 387382R6 (THYMNNOT02)	17 1937	1291 1999 1219 1887
				5262448T6 (CONDTUT02)	1824	
				71193135V1 71198067V1 71191147V1 71189347V1 71391065V1	539 2173 421 1291 1219	

Table 4 (cont.)

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
19	7182343CB1	2178	1-682, 1280-1431	70647146V1 7182637H1 (BONRFE01)	1100 1155	1778 1801
				7358741H1 7182343H1 (BONRFE01) 6951475H1 (BRAITDR02)	613 1 193	1156 503 807
				706760886V1	1622	2178
20	7264345CB1	2777	1710-2038, 651-1227, 354-393	7264345H1 (PROSTM02) 70003406D1 6609118H2 (PLACFEC01) 70003129D1	580 708 1 2280	1154 1372 597 2777
				711113886V1 7014503H1 (KIDNNOC01) 70384792D1	1813 1988 1183	2512 2547 1702
				70006067D1	1360	1908
				GNN.g6249409_000023_002	1	972
21	7472328CB1	972	444-467	135966T6 (BMARNOT02)	1820	2514
22	1471412CB1	2538	547-1835 1-113	70162584V1 70604897V1 2489917R6 (BRAVUNT02)	1085 1001 1	1758 1650 468
				5504821R6 (BRADDIRO1) 4741826R8 (THYMNOR02) 70608907V1	363 483 2019	996 1073 2538
				70162579V1	1687	2311

**Table 5**

Polynucleotide SEQ ID NO:	Incyte Project ID	Representative Library
12	4912053CB1	UCMCNOT02
13	870615CB1	BRABDIR01
14	2762870CB1	BRSTTUT03
15	5507012CB1	BRAUNOR01
16	6756529CB1	BLADDNOT09
17	387382CB1	THYMNNOT02
18	5850786CB1	CONDTUT02
19	7182343CB1	FTUBTUT01
20	7264345CB1	EOS1HET02
22	1471412CB1	PROSTUS23

Table 6

Library	Vector	Library Description
BRABDI01	PINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident. Patient history included Huntington's disease, emphysema, and tobacco abuse.
UCMCNOT02	PINCY	Library was constructed using RNA isolated from mononuclear cells obtained from the umbilical cord blood of nine individuals.
BRSTTUT03	PSPORT1	Library was constructed using RNA isolated from breast tumor tissue removed from a 58-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology indicated multicentric invasive grade 4 lobular carcinoma. The mass was identified in the upper outer quadrant, and three separate nodules were found in the lower outer quadrant of the left breast.
BLADNOT09	PINCY	Library was constructed using RNA isolated from bladder tissue removed from a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, regional lymph node excision, and urinary diversion to bowel. Pathology for the associated tumor tissue indicated invasive grade 3 transitional cell carcinoma. Patient history included a benign colon neoplasm.
BRAUNOR01	PINCY	This random primed library was constructed using RNA isolated from striatum, globus pallidus and posterior putamen tissue removed from an 81-year-old Caucasian female who died from a hemorrhage and ruptured thoracic aorta due to atherosclerosis. Pathology indicated moderate atherosclerosis involving the internal carotids, bilaterally; microscopic infarcts of the frontal cortex and hippocampus; and scattered diffuse amyloid plaques and neurofibrillary tangles, consistent with age. Grossly, the leptomeninges showed only mild thickening and hyalinization along the superior sagittal sinus. The remainder of the leptomeninges was thin and contained some congested blood vessels. Mild atrophy was found mostly in the frontal poles and lobes, and temporal lobes, bilaterally. Microscopically, there were pairs of Alzheimer type II astrocytes within the deep layers of the neocortex. There was increased sateillitosis around neurons in the deep gray matter in the middle frontal cortex. The amygdala contained rare diffuse plaques and neurofibrillary tangles. The posterior hippocampus contained a microscopic area of cystic cavitation with hemosiderin-laden macrophages surrounded by reactive gliosis. Patient history included sepsis, cholangitis, post-operative atelectasis, pneumonia CAD, cardiomegaly due to left ventricular hypertrophy, splenomegaly, arteriolonephrosclerosis, nodular colloid goiter, emphysema, CHF, hypothyroidism, and peripheral vascular disease.

Table 6 (cont.)

Library	Vector	Library Description
CONDUT02	PINCY	Library was constructed using RNA isolated from peritoneal tumor tissue removed from a 66-year-old Caucasian female during exploratory laparotomy. Pathology indicated neuroendocrine carcinoma involving the anterior peritoneum, right ovary and fallopian tube, appendix, omentum, left diaphragm, and the small bowel, sigmoid, and uterine serosa. The uterus contained an endometrial polyp and calcified leiomyoma. Multiple lymph nodes were positive for tumor. Patient history included melanosis coli.
ESTHET02	PBLUESCRIPT	The library was constructed using RNA isolated from peripheral blood cells apheresed from a 48-year-old Caucasian male. Patient history included hyperesinophilia. The cell population was determined to be greater than 77% eosinophils by Wright's staining.
FTUBTUR01	PCDNA2.1	This random primed library was constructed using RNA isolated from fallopian tube tumor tissue removed from an 85-year-old Caucasian female during bilateral salpingo-oophorectomy and hysterectomy. Pathology indicated poorly differentiated mixed endometrioid (80%) and serous (20%) adenocarcinoma, which was confined to the mucosa without mural involvement. Endometrioid carcinoma in situ was also present. Pathology for the associated uterus tumor indicated focal endometrioid adenocarcinoma in situ and moderately differentiated invasive adenocarcinoma arising in an endometrial polyp. Metastatic endometrioid and serous adenocarcinoma was present at the cul-de-sac tumor. Patient history included medullary carcinoma of the thyroid and myocardial infarction.
THYMNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from thymus tissue removed from a 3-year-old Caucasian male, who died from drowning.
PROSTUSS23	PINCY	This subtracted prostate tumor library was constructed using 1 million clones from a pooled prostate tumor library that was subjected to 2 rounds of subtractive hybridization with 1 million clones from a pooled prostate tissue library. The starting library for subtraction was constructed by pooling equal numbers of clones from 4 prostate tumor libraries using mRNA isolated from prostate tumor removed from Caucasian males at ages 58 (A), 61 (B), 66 (C), and 68 (D) during prostatectomy with lymph node excision. Pathology indicated adenoca in all donors. History included elevated PSA, induration and tobacco abuse in donor A; elevated PSA, induration, prostate hyperplasia, renal failure, osteoarthritis, renal artery stenosis, benign HTN, thrombocytopenia, hyperlipidemia, tobacco/alcohol abuse and hepatitis C (carrier) in donor B; elevated PSA, induration, and tobacco abuse in donor C; and elevated PSA, induration, hypercholesterolemia, and kidney calculus in donor D. The hybridization probe for subtraction was constructed by pooling equal numbers of cDNA clones from 3 prostate tissue libraries derived from prostate tissue, prostate epithelial cells, and fibroblasts from prostate stroma from 3 different donors. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR 19 (1991):1954 and Bonaldo, et al. Genome Research 6 (1996):791.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABIPARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	<i>ESTs</i> : Probability value= 1.0E-8 or less <i>Full Length sequences</i> : Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>ESTs</i> : fasta E value=1.06E-6 <i>Assembled ESTs</i> : fasta Identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	<i>PFAM hits</i> : Probability value= 1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

**Table 7 (cont.)**

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phil's Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score= 120 or greater; Match length= 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of:
  - 5 a) an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,
  - b) a naturally occurring amino acid sequence having at least 90% sequence identity to an amino acid sequence selected from the group consisting of SEQ ID NO:1-11,
  - c) a biologically active fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11, and
  - 10 d) an immunogenic fragment of an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.
2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-11.
- 15 3. An isolated polynucleotide encoding a polypeptide of claim 1.
4. An isolated polynucleotide encoding a polypeptide of claim 2.
- 20 5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:12-22.
6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
- 25 7. A cell transformed with a recombinant polynucleotide of claim 6.
8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
- 30 9. A method for producing a polypeptide of claim 1, the method comprising:
  - a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

b) recovering the polypeptide so expressed.

10. An isolated antibody which specifically binds to a polypeptide of claim 1.

5 11. An isolated polynucleotide comprising a polynucleotide sequence selected from the group consisting of:

- a) a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,
- b) a naturally occurring polynucleotide sequence having at least 90% sequence identity to a polynucleotide sequence selected from the group consisting of SEQ ID NO:12-22,
- 10 c) a polynucleotide sequence complementary to a),
- d) a polynucleotide sequence complementary to b), and
- e) an RNA equivalent of a)-d).

12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a  
15 polynucleotide of claim 11.

13. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

20 a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and  
b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.

25

14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

30 a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and  
b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising an effective amount of a polypeptide of claim 1 and a pharmaceutically acceptable excipient.
17. A composition of claim 16, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-11.
  18. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment the composition of claim 16.
19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:
  - a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
  - b) detecting agonist activity in the sample.
20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.
21. A method for treating a disease or condition associated with decreased expression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 20.
22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:
  - a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
  - b) detecting antagonist activity in the sample.
23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.
24. A method for treating a disease or condition associated with overexpression of functional PKIN, comprising administering to a patient in need of such treatment a composition of claim 23.
25. A method of screening for a compound that specifically binds to the polypeptide of claim

1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;
- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the

amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

<110> INCYTE GENOMICS, INC.  
TANG, Y. Tom  
BURFORD, Neil  
GHANDI, Ameena R.  
PATTERSON, Chandra  
KHAN, Farrah A.  
YUE, Henry  
HAFALIA, April  
SHIH, Leo L.  
TRIBOULEY, Catherine M.  
YAO, Monique G.  
BURRILL, John D.  
MARCUS, Gregory A.  
ZINGLER, Kurt A.  
LU, Dyung Aina M.  
BANDMAN, Olga  
POLICKY, Jennifer L.  
GRIFFIN, Jennifer A.  
THORNTON, Michael  
NGUYEN, Dannie B.  
LAL, Preeti  
WALSH, Roderick T.

<120> HUMAN KINASES

<130> PI-0037 PCT

<140> To Be Assigned  
<141> Herewith

<160> 22

<170> PERL Program

<210> 1

<211> 792

<212> PRT

<213> Homo sapiens

<220>

<221> misc\_feature

<223> Incyte ID No: 4912053CD1

<400> 1

Met	Lys	Glu	His	Phe	Phe	Ser	Ala	Ser	Phe	Ser	Asp	Thr	Lys	Leu
1				5					10				15	
Lys	Pro	Gly	Thr	Val	Ser	Glu	Gly	Leu	Ser	Phe	Leu	His	Ser	Ser
					20				25				30	
Val	Lys	Met	Val	His	Gly	Asn	Ile	Thr	Pro	Glu	Asn	Ile	Ile	Leu
					35				40				45	
Asn	Lys	Ser	Gly	Ala	Trp	Lys	Ile	Met	Gly	Phe	Asp	Phe	Cys	Val
					50				55				60	
Ser	Ser	Thr	Asn	Pro	Ser	Glu	Gln	Glu	Pro	Lys	Phe	Pro	Cys	Lys
					65				70				75	
Glu	Trp	Asp	Pro	Asn	Leu	Pro	Ser	Leu	Cys	Leu	Pro	Asn	Pro	Glu
					80				85				90	
Tyr	Leu	Ala	Pro	Glu	Tyr	Ile	Leu	Ser	Val	Ser	Cys	Glu	Thr	Ala
					95				100				105	
Ser	Asp	Met	Tyr	Ser	Leu	Gly	Thr	Val	Met	Tyr	Ala	Val	Phe	Asn
					110				115				120	
Lys	Gly	Lys	Pro	Ile	Phe	Glu	Val	Asn	Lys	Gln	Asp	Ile	Tyr	Lys
					125				130				135	
Ser	Phe	Ser	Arg	Gln	Leu	Asp	Gln	Leu	Ser	Arg	Leu	Gly	Ser	Ser
					140				145				150	

Ser Leu Thr Asn Ile Pro Glu Glu Val Arg Glu His Val Lys Leu  
                   155                  160                  165  
 Leu Leu Asn Val Thr Pro Thr Val Arg Pro Asp Ala Asp Gln Met  
                   170                  175                  180  
 Thr Lys Ile Pro Phe Phe Asp Asp Val Gly Ala Val Thr Leu Gln  
                   185                  190                  195  
 Tyr Phe Asp Thr Leu Phe Gln Arg Asp Asn Leu Gln Lys Ser Gln  
                   200                  205                  210  
 Phe Phe Lys Gly Leu Pro Lys Val Leu Pro Lys Leu Pro Lys Arg  
                   215                  220                  225  
 Val Ile Val Gln Arg Ile Leu Pro Cys Leu Thr Ser Glu Phe Val  
                   230                  235                  240  
 Asn Pro Asp Met Val Pro Phe Val Leu Pro Asn Val Leu Leu Ile  
                   245                  250                  255  
 Ala Glu Glu Cys Thr Lys Glu Glu Tyr Val Lys Leu Ile Leu Pro  
                   260                  265                  270  
 Glu Leu Gly Pro Val Phe Lys Gln Gln Glu Pro Ile Gln Ala Ser  
                   275                  280                  285  
 Asn Met Ile Leu Leu Ile Phe Leu Gln Lys Met Asp Leu Leu Leu  
                   290                  295                  300  
 Thr Lys Thr Pro Pro Asp Glu Ile Lys Asn Ser Val Leu Pro Met  
                   305                  310                  315  
 Val Tyr Arg Ala Leu Glu Ala Pro Ser Ile Gln Ile Gln Glu Leu  
                   320                  325                  330  
 Cys Leu Asn Ile Ile Pro Thr Phe Ala Asn Leu Ile Asp Tyr Pro  
                   335                  340                  345  
 Ser Met Lys Asn Ala Leu Ile Pro Arg Ile Lys Asn Ala Cys Leu  
                   350                  355                  360  
 Gln Thr Ser Ser Leu Ala Val Arg Val Asn Ser Leu Val Cys Leu  
                   365                  370                  375  
 Gly Lys Ile Leu Glu Tyr Leu Asp Lys Trp Phe Val Leu Asp Asp  
                   380                  385                  390  
 Ile Leu Pro Phe Leu Gln Gln Ile Pro Ser Lys Glu Pro Ala Val  
                   395                  400                  405  
 Leu Met Gly Ile Leu Gly Ile Tyr Lys Cys Thr Phe Thr His Lys  
                   410                  415                  420  
 Lys Leu Gly Ile Thr Lys Glu Gln Leu Ala Gly Lys Val Leu Pro  
                   425                  430                  435  
 His Leu Ile Pro Leu Ser Ile Glu Asn Asn Leu Asn Leu Asn Gln  
                   440                  445                  450  
 Phe Asn Ser Phe Ile Ser Val Ile Lys Glu Met Leu Asn Arg Leu  
                   455                  460                  465  
 Glu Ser Glu His Lys Thr Lys Leu Glu Gln Leu His Ile Met Gln  
                   470                  475                  480  
 Glu Gln Gln Lys Ser Leu Asp Ile Gly Asn Gln Met Asn Val Ser  
                   485                  490                  495  
 Glu Glu Met Lys Val Thr Asn Ile Gly Asn Gln Gln Ile Asp Lys  
                   500                  505                  510  
 Val Phe Asn Asn Ile Gly Ala Asp Leu Leu Thr Gly Ser Glu Ser  
                   515                  520                  525  
 Glu Asn Lys Glu Asp Gly Leu Gln Asn Lys His Lys Arg Ala Ser  
                   530                  535                  540  
 Leu Thr Leu Glu Glu Lys Gln Lys Leu Ala Lys Glu Gln Glu Gln  
                   545                  550                  555  
 Ala Gln Lys Leu Lys Ser Gln Gln Pro Leu Lys Pro Gln Val His  
                   560                  565                  570  
 Thr Pro Val Ala Thr Val Lys Gln Thr Lys Asp Leu Thr Asp Thr  
                   575                  580                  585  
 Leu Met Asp Asn Met Ser Ser Leu Thr Ser Leu Ser Val Ser Thr  
                   590                  595                  600  
 Pro Lys Ser Ser Ala Ser Ser Thr Phe Thr Ser Val Pro Ser Met  
                   605                  610                  615  
 Gly Ile Gly Met Met Phe Ser Thr Pro Thr Asp Asn Thr Lys Arg

	620		625		630									
Asn	Leu	Thr	Asn	Gly	Leu	Asn	Ala	Asn	Met	Gly	Phe	Gln	Thr	Ser
				635					640				645	
Gly	Phe	Asn	Met	Pro	Val	Asn	Thr	Asn	Gln	Asn	Phe	Tyr	Ser	Ser
				650					655				660	
Pro	Ser	Thr	Val	Gly	Val	Thr	Lys	Met	Thr	Leu	Gly	Thr	Pro	Pro
				665					670				675	
Thr	Leu	Pro	Asn	Phe	Asn	Ala	Leu	Ser	Val	Pro	Pro	Ala	Gly	Ala
				680					685				690	
Lys	Gln	Thr	Gln	Gln	Arg	Pro	Thr	Asp	Met	Ser	Ala	Leu	Asn	Asn
				695					700				705	
Leu	Phe	Gly	Pro	Gln	Lys	Pro	Lys	Val	Ser	Met	Asn	Gln	Leu	Ser
				710					715				720	
Gln	Gln	Lys	Pro	Asn	Gln	Trp	Leu	Asn	Gln	Phe	Val	Pro	Pro	Gln
				725					730				735	
Gly	Ser	Pro	Thr	Met	Gly	Ser	Ser	Val	Met	Gly	Thr	Gln	Met	Asn
				740					745				750	
Val	Ile	Gly	Gln	Ser	Ala	Phe	Gly	Met	Gln	Gly	Asn	Pro	Phe	Phe
				755					760				765	
Asn	Pro	Gln	Asn	Phe	Ala	Gln	Pro	Pro	Thr	Thr	Met	Thr	Asn	Ser
				770					775				780	
Ser	Ser	Ala	Ser	Asn	Asp	Leu	Lys	Asp	Leu	Phe	Gly			
				785					790					

&lt;210&gt; 2

&lt;211&gt; 224

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 870615CD1

&lt;400&gt; 2

Met	Gly	Gly	Gly	Gly	Gly	Arg	Gly	Ser	Gly	Gly	Gly	Met	Asp	Gly
1					5				10					15
Trp	Asp	Gly	Ser	Ile	Arg	Glu	Val	Leu	Tyr	Arg	Gly	Asn	Ala	Asp
				20					25					30
Glu	Glu	Leu	Phe	Gln	Asp	Gln	Asp	Asp	Asp	Glu	Leu	Trp	Leu	Gly
				35					40					45
Asp	Ser	Gly	Ile	Thr	Asn	Trp	Asp	Asn	Val	Asp	Tyr	Met	Trp	Asp
				50					55					60
Glu	Asp	Gln	Asp	Tyr	Tyr	Leu	Gly	Gly						
				65					70					75
Leu	Arg	Pro	Asp	Leu	Arg	Ile	Asp	Val	Tyr	Arg	Glu	Glu	Ile	
				80					85					90
Leu	Glu	Ala	Tyr	Asp	Glu	Asp	Glu	Asp	Glu	Glu	Leu	Tyr	Pro	Asp
				95					100					105
Ile	His	Pro	Pro	Pro	Ser	Leu	Pro	Leu	Pro	Gly	Gln	Phe	Thr	Cys
				110					115					120
Pro	Gln	Cys	Arg	Lys	Ser	Phe	Thr	Arg	Arg	Ser	Phe	Arg	Pro	Asn
				125					130					135
Leu	Gln	Leu	Ala	Asn	Met	Val	Gln	Ile	Ile	Arg	Gln	Met	Cys	Pro
				140					145					150
Thr	Pro	Tyr	Arg	Gly	Asn	Arg	Ser	Asn	Asp	Gln	Gly	Met	Cys	Phe
				155					160					165
Lys	His	Gln	Glu	Ala	Leu	Lys	Leu	Phe	Cys	Glu	Val	Asp	Lys	Glu
				170					175					180
Ala	Ile	Cys	Val	Val	Cys	Arg	Glu	Ser	Arg	Ser	His	Lys	Gln	His
				185					190					195
Ser	Val	Val	Pro	Leu	Glu	Glu	Val	Val	Gln	Glu	Tyr	Gln	Lys	Ile
				200					205					210
Gly	Ser	Thr	Ser	Ser	Val	Glu	Leu	Ser	Glu	Ser	Val	Gly	Gln	

215

220

<210> 3  
<211> 374  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 2762870CD1

<400> 3

Met	Ala	Ala	Pro	Ala	Leu	Gly	Leu	Val	Cys	Gly	Arg	Cys	Pro	Glu
1				5					10					15
Leu	Gly	Leu	Val	Leu	Leu	Leu	Leu	Leu	Ser	Leu	Leu	Cys	Gly	
					20				25					30
Ala	Ala	Gly	Ser	Gln	Glu	Ala	Gly	Thr	Gly	Ala	Gly	Ala	Gly	Ser
				35					40					45
Leu	Ala	Gly	Ser	Cys	Gly	Cys	Gly	Thr	Pro	Gln	Arg	Pro	Gly	Ala
				50					55					60
His	Gly	Ser	Ser	Ala	Ala	Ala	His	Arg	Tyr	Ser	Arg	Glu	Ala	Asn
				65					70					75
Ala	Pro	Gly	Pro	Val	Pro	Gly	Glu	Arg	Gln	Leu	Ala	His	Ser	Lys
				80					85					90
Met	Val	Pro	Ile	Pro	Ala	Gly	Val	Phe	Thr	Met	Gly	Thr	Asp	Asp
				95					100					105
Pro	Gln	Ile	Lys	Gln	Asp	Gly	Glu	Ala	Pro	Ala	Arg	Arg	Val	Thr
				110					115					120
Ile	Asp	Ala	Phe	Tyr	Met	Asp	Ala	Tyr	Glu	Val	Ser	Asn	Thr	Glu
				125					130					135
Phe	Glu	Lys	Phe	Val	Asn	Ser	Thr	Gly	Tyr	Leu	Thr	Glu	Ala	Glu
				140					145					150
Lys	Phe	Gly	Asp	Ser	Phe	Val	Phe	Glu	Gly	Met	Leu	Ser	Glu	Gln
				155					160					165
Val	Lys	Thr	Asn	Ile	Gln	Gln	Ala	Val	Ala	Ala	Ala	Pro	Trp	Trp
				170					175					180
Leu	Pro	Val	Lys	Gly	Ala	Asn	Trp	Arg	His	Pro	Glu	Gly	Pro	Asp
				185					190					195
Ser	Thr	Ile	Leu	His	Arg	Pro	Asp	His	Pro	Val	Leu	His	Val	Ser
				200					205					210
Trp	Asn	Asp	Ala	Val	Ala	Tyr	Cys	Thr	Trp	Ala	Gly	Lys	Arg	Leu
				215					220					225
Pro	Thr	Glu	Ala	Glu	Trp	Glu	Tyr	Ser	Cys	Arg	Gly	Gly	Leu	His
				230					235					240
Asn	Arg	Leu	Phe	Pro	Trp	Gly	Asn	Lys	Leu	Gln	Pro	Lys	Gly	Gln
				245					250					255
His	Tyr	Ala	Asn	Ile	Trp	Gln	Gly	Glu	Phe	Pro	Val	Thr	Asn	Thr
				260					265					270
Gly	Glu	Asp	Gly	Phe	Gln	Gly	Thr	Ala	Pro	Val	Asp	Ala	Phe	Pro
				275					280					285
Pro	Asn	Gly	Tyr	Gly	Leu	Tyr	Asn	Ile	Val	Gly	Asn	Ala	Trp	Glu
				290					295					300
Trp	Thr	Ser	Asp	Trp	Trp	Thr	Val	His	His	Ser	Val	Glu	Glu	Thr
				305					310					315
Leu	Asn	Pro	Lys	Gly	Pro	Pro	Ser	Gly	Lys	Asp	Arg	Val	Lys	Lys
				320					325					330
Gly	Gly	Ser	Tyr	Met	Cys	His	Arg	Ser	Tyr	Cys	Tyr	Arg	Tyr	Arg
				335					340					345
Cys	Ala	Ala	Arg	Ser	Gln	Asn	Thr	Pro	Asp	Ser	Ser	Ala	Ser	Asn
				350					355					360
Leu	Gly	Phe	Arg	Cys	Ala	Ala	Asp	Arg	Leu	Pro	Thr	Met	Asp	
				365					370					

<210> 4  
<211> 565  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5507012CD1

<400> 4

Met	Ser	Gly	Gly	Ala	Ala	Glu	Lys	Gln	Ser	Ser	Thr	Pro	Gly	Ser
1				5					10					15
Leu	Phe	Leu	Ser	Pro	Pro	Ala	Pro	Ala	Pro	Lys	Asn	Gly	Ser	Ser
					20				25					30
Ser	Asp	Ser	Ser	Val	Gly	Glu	Lys	Leu	Gly	Ala	Ala	Ala	Ala	Asp
					35				40					45
Ala	Val	Thr	Gly	Arg	Thr	Glu	Glu	Tyr	Arg	Arg	Arg	Arg	His	Thr
					50				55					60
Met	Asp	Lys	Asp	Ser	Arg	Gly	Ala	Ala	Ala	Thr	Thr	Thr	Thr	Thr
					65				70					75
Glu	His	Arg	Phe	Phe	Arg	Arg	Ser	Val	Ile	Cys	Asp	Ser	Asn	Ala
					80				85					90
Thr	Ala	Leu	Glu	Leu	Pro	Gly	Leu	Pro	Leu	Ser	Leu	Pro	Gln	Pro
					95				100					105
Ser	Ile	Pro	Ala	Ala	Val	Pro	Gln	Ser	Ala	Pro	Pro	Glu	Pro	His
					110				115					120
Arg	Glu	Glu	Thr	Val	Thr	Ala	Thr	Ala	Thr	Ser	Gln	Val	Ala	Gln
					125				130					135
Gln	Pro	Pro	Ala	Ala	Ala	Ala	Pro	Gly	Glu	Gln	Ala	Val	Ala	Gly
					140				145					150
Pro	Ala	Pro	Ser	Thr	Val	Pro	Ser	Ser	Thr	Ser	Lys	Asp	Arg	Pro
					155				160					165
Val	Ser	Gln	Pro	Ser	Leu	Val	Gly	Ser	Lys	Glu	Glu	Pro	Pro	Pro
					170				175					180
Ala	Arg	Ser	Gly	Ser	Gly	Gly	Gly	Ser	Ala	Lys	Glu	Pro	Gln	Glu
					185				190					195
Glu	Arg	Ser	Gln	Gln	Gln	Asp	Asp	Ile	Glu	Glu	Leu	Glu	Thr	Lys
					200				205					210
Ala	Val	Gly	Met	Ser	Asn	Asp	Gly	Arg	Phe	Leu	Lys	Phe	Asp	Ile
					215				220					225
Glu	Ile	Gly	Arg	Gly	Ser	Phe	Lys	Thr	Val	Tyr	Lys	Gly	Leu	Asp
					230				235					240
Thr	Glu	Thr	Thr	Val	Glu	Val	Ala	Trp	Cys	Glu	Leu	Gln	Asp	Arg
					245				250					255
Lys	Leu	Thr	Lys	Ser	Glu	Arg	Gln	Arg	Phe	Lys	Glu	Glu	Ala	Glu
					260				265					270
Met	Leu	Lys	Gly	Leu	Gln	His	Pro	Asn	Ile	Val	Arg	Phe	Tyr	Asp
					275				280					285
Ser	Trp	Glu	Ser	Thr	Val	Lys	Gly	Lys	Lys	Cys	Ile	Val	Leu	Val
					290				295					300
Thr	Glu	Leu	Met	Thr	Ser	Gly	Thr	Leu	Lys	Thr	Tyr	Leu	Lys	Arg
					305				310					315
Phe	Lys	Val	Met	Lys	Ile	Lys	Val	Leu	Arg	Ser	Trp	Cys	Arg	Gln
					320				325					330
Ile	Leu	Lys	Gly	Leu	Gln	Phe	Leu	His	Thr	Arg	Thr	Pro	Pro	Ile
					335				340					345
Ile	His	Arg	Asp	Leu	Lys	Cys	Asp	Asn	Ile	Phe	Ile	Thr	Gly	Pro
					350				355					360
Thr	Gly	Ser	Val	Lys	Ile	Gly	Asp	Leu	Gly	Leu	Ala	Thr	Leu	Lys
					365				370					375
Arg	Ala	Ser	Phe	Ala	Lys	Ser	Val	Ile	Gly	Thr	Pro	Glu	Phe	Met
					380				385					390
Ala	Pro	Glu	Met	Tyr	Glu	Glu	Lys	Tyr	Asp	Glu	Ser	Val	Asp	Val

	395		400		405
Tyr Ala Phe Gly	Met Cys Met Leu Glu		Met Ala Thr Ser Glu	Tyr	
	410		415		420
Pro Tyr Ser Glu	Cys Gln Asn Ala Ala	Gln	Ile Tyr Arg Arg	Val	
	425		430		435
Thr Ser Gly Val	Lys Pro Ala Ser Phe	Asp	Lys Val Ala Ile	Pro	
	440		445		450
Glu Val Lys Glu	Ile Ile Glu Gly Cys	Ile	Arg Gln Asn Lys	Asp	
	455		460		465
Glu Arg Tyr Ser	Ile Lys Asp Leu Leu	Asn	His Ala Phe Phe	Gln	
	470		475		480
Glu Glu Thr Gly	Val Arg Val Glu Leu	Ala	Glu Glu Asp Asp	Gly	
	485		490		495
Glu Lys Ile Ala	Ile Lys Leu Trp Leu	Arg	Ile Glu Asp Ile	Lys	
	500		505		510
Lys Leu Lys Gly	Lys Tyr Lys Asp Asn	Glu	Ala Ile Glu Phe	Ser	
	515		520		525
Phe Asp Leu Glu	Arg Asp Val Pro Glu	Asp	Val Ala Gln Glu	Met	
	530		535		540
Val Glu Ser Gly	Tyr Val Cys Glu Gly	Asp	His Lys Thr Met	Ala	
	545		550		555
Lys Ala Phe Lys	Asp Arg Val Ser Leu	Ile			
	560		565		

&lt;210&gt; 5

&lt;211&gt; 783

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 6756529CD1

&lt;400&gt; 5

Met Val Ile Met Ser Glu Phe Ser Ala Asp	Pro Ala Gly Gln Gly		
1	5	10	15
Gln Gly Gln Gln Lys Pro Leu Arg Val	Gly Phe Tyr Asp Ile Glu		
	20	25	30
Arg Thr Leu Gly Lys Gly Asn Phe Ala Val	Val Val Lys Leu Ala Arg		
	35	40	45
His Arg Val Thr Lys Thr Gln Val Ala Ile	Lys Ile Ile Asp Lys		
	50	55	60
Thr Arg Leu Asp Ser Ser Asn Leu Glu	Lys Ile Tyr Arg Glu Val		
	65	70	75
Gln Leu Met Lys Leu Leu Asn His Pro	His Ile Ile Lys Leu Tyr		
	80	85	90
Gln Val Met Glu Thr Lys Asp Met Leu	Tyr Ile Val Thr Glu Phe		
	95	100	105
Ala Lys Asn Gly Glu Met Phe Asp Tyr	Leu Thr Ser Asn Gly His		
	110	115	120
Leu Ser Glu Asn Glu Ala Arg Lys Lys	Phe Trp Gln Ile Leu Ser		
	125	130	135
Ala Val Glu Tyr Cys His Asp His His	Ile Val His Arg Asp Leu		
	140	145	150
Lys Thr Glu Asn Leu Leu Asp Gly	Asn Met Asp Ile Lys Leu		
	155	160	165
Ala Asp Phe Gly Phe Gly Asn Phe Tyr	Lys Ser Gly Glu Pro Leu		
	170	175	180
Ser Thr Trp Cys Gly Ser Pro Pro Tyr	Ala Ala Pro Glu Val Phe		
	185	190	195
Glu Gly Lys Glu Tyr Glu Gly Pro Gln	Leu Asp Ile Trp Ser Leu		
	200	205	210
Gly Val Val Leu Tyr Val Leu Val Cys	Gly Ser Leu Pro Phe Asp		

	215	220	225
Gly Pro Asn Leu Pro	Thr Leu Arg Gln Arg	Val Leu Glu Gly	Arg
230	235	240	
Phe Arg Ile Pro Phe	Phe Met Ser Gln Asp	Cys Glu Ser Leu	Ile
245	250	255	
Arg Arg Met Leu Val	Val Asp Pro Ala Arg	Arg Ile Thr Ile	Ala
260	265	270	
Gln Ile Arg Gln His	Arg Trp Met Arg Ala	Glu Pro Cys Leu	Pro
275	280	285	
Gly Pro Ala Cys Pro	Ala Phe Ser Ala His	Ser Tyr Thr Ser	Asn
290	295	300	
Leu Gly Asp Tyr Asp	Glu Gln Ala Leu Gly	Ile Met Gln Thr	Leu
305	310	315	
Gly Val Asp Arg Gln	Arg Thr Val Glu Ser	Leu Gln Asn Ser	Ser
320	325	330	
Tyr Asn His Phe Ala	Ala Ile Tyr Tyr Leu	Leu Leu Glu Arg	Leu
335	340	345	
Lys Glu Tyr Arg Asn	Ala Gln Cys Ala Arg	Pro Gly Pro Ala	Arg
350	355	360	
Gln Pro Arg Pro Arg	Ser Ser Asp Leu Ser	Gly Leu Glu Val	Pro
365	370	375	
Gln Glu Gly Leu Ser	Thr Asp Pro Phe Arg	Pro Ala Leu Leu	Cys
380	385	390	
Pro Gln Pro Gln Thr	Leu Val Gln Ser Val	Leu Gln Ala Glu	Met
395	400	405	
Asp Cys Glu Leu Gln	Ser Ser Leu Gln Trp	Pro Leu Phe Phe	Pro
410	415	420	
Val Asp Ala Ser Cys	Ser Gly Val Phe Arg	Pro Arg Pro Val	Ser
425	430	435	
Pro Ser Ser Leu Leu	Asp Thr Ala Ile Ser	Glu Glu Ala Arg	Gln
440	445	450	
Gly Pro Gly Leu Glu	Glu Glu Gln Asp Thr	Gln Glu Ser Leu	Pro
455	460	465	
Ser Ser Thr Gly Arg	Arg Arg His Thr Leu Ala	Glu Val Ser Thr	Arg
470	475	480	
Leu Ser Pro Leu Thr	Ala Pro Cys Ile Val	Val Ser Pro Ser	Thr
485	490	495	
Thr Ala Ser Pro Ala	Glu Gly Thr Ser Ser	Asp Ser Cys Leu	Thr
500	505	510	
Phe Ser Ala Ser Lys	Ser Pro Ala Gly Leu	Ser Gly Thr Pro	Ala
515	520	525	
Thr Gln Gly Leu Leu	Gly Ala Cys Ser Pro	Val Arg Leu Ala	Ser
530	535	540	
Pro Phe Leu Gly Ser	Gln Ser Ala Thr Pro	Val Leu Gln Ala	Gln
545	550	555	
Gly Gly Leu Gly Gly	Ala Val Leu Leu Pro	Val Ser Phe Gln	Glu
560	565	570	
Gly Arg Arg Ala Ser	Asp Thr Ser Leu Thr	Gln Gly Leu Lys	Ala
575	580	585	
Phe Arg Gln Gln Leu	Arg Lys Thr Arg Thr	Lys Gly Phe Leu	
590	595	600	
Gly Leu Asn Lys Ile	Lys Gly Leu Ala Arg	Gln Val Cys Gln	Val
605	610	615	
Pro Ala Ser Arg Ala	Ser Arg Gly Gly Leu	Ser Pro Phe His	Ala
620	625	630	
Pro Ala Gln Ser Pro	Gly Leu His Gly Gly	Ala Gly Gly Ser	Arg
635	640	645	
Glu Gly Trp Ser Leu	Leu Glu Glu Val Leu	Glu Gln Gln Arg	Leu
650	655	660	
Leu Gln Leu Gln His	His Pro Ala Ala Ala	Pro Gly Cys Ser	Gln
665	670	675	
Ala Pro Gln Pro Ala	Pro Ala Pro Phe Val	Ile Ala Pro Cys	Asp
680	685	690	

Gly Pro Gly Ala Ala Pro Leu Pro Ser Thr Leu Leu Thr Ser Gly  
                   695                 700                 705  
 Leu Pro Leu Leu Pro Pro Pro Leu Leu Gln Thr Gly Ala Ser Pro  
                   710                 715                 720  
 Val Ala Ser Ala Ala Gln Leu Leu Asp Thr His Leu His Ile Gly  
                   725                 730                 735  
 Thr Gly Pro Thr Ala Leu Pro Ala Val Pro Pro Pro Arg Leu Ala  
                   740                 745                 750  
 Arg Leu Ala Pro Gly Cys Glu Pro Leu Gly Leu Leu Gln Gly Asp  
                   755                 760                 765  
 Cys Glu Met Glu Asp Leu Met Pro Cys Ser Leu Gly Thr Phe Val  
                   770                 775                 780  
 Leu Val Gln

<210> 6  
<211> 798  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 387382CD1

<400> 6  
 Met Ala Ala Ser Leu Pro Gly Pro Gly Ser Arg Leu Phe His Thr  
     1              5                 10                 15  
 Tyr Gly Ala Ala Asp Gly Arg Arg Gln Arg Arg Pro Gly Arg Glu  
     20                 25                 30  
 Ala Ala Gln Trp Phe Pro Pro Gln Asp Arg Arg Arg Phe Phe Asn  
     35                 40                 45  
 Ser Ser Gly Ser Ser Asp Ala Ser Ile Gly Asp Pro Ser Gln Ser  
     50                 55                 60  
 Asp Asp Pro Asp Asp Pro Asp Asp Pro Asp Phe Pro Gly Ser Pro  
     65                 70                 75  
 Val Arg Arg Arg Arg Arg Cys Pro Gly Gly Arg Val Pro Lys Asp  
     80                 85                 90  
 Arg Pro Ser Leu Thr Val Thr Pro Lys Arg Trp Lys Leu Arg Ala  
     95                 100                105  
 Arg Pro Ser Leu Thr Val Thr Pro Arg Arg Leu Gly Leu Arg Ala  
     110                 115                120  
 Arg Pro Pro Gln Lys Cys Ser Thr Pro Cys Gly Pro Leu Arg Leu  
     125                 130                135  
 Pro Pro Phe Pro Ser Arg Asp Ser Gly Arg Leu Ser Pro Asp Leu  
     140                 145                150  
 Ser Val Cys Gly Gln Pro Arg Asp Gly Asp Glu Leu Gly Ile Ser  
     155                 160                165  
 Ala Ser Leu Phe Ser Ser Leu Ala Ser Pro Cys Pro Gly Ser Pro  
     170                 175                180  
 Thr Pro Arg Asp Ser Val Ile Ser Ile Gly Thr Ser Ala Cys Leu  
     185                 190                195  
 Val Ala Ala Ser Ala Val Pro Ser Gly Leu His Leu Pro Glu Val  
     200                 205                210  
 Ser Leu Asp Arg Ala Ser Leu Pro Cys Ser Gln Glu Glu Ala Thr  
     215                 220                225  
 Gly Gly Ala Lys Asp Thr Arg Met Val His Gln Thr Arg Ala Ser  
     230                 235                240  
 Leu Arg Ser Val Leu Phe Gly Leu Met Asn Ser Gly Thr Pro Glu  
     245                 250                255  
 Asp Ser Glu Phe Arg Ala Asp Gly Lys Asn Met Arg Glu Ser Cys  
     260                 265                270  
 Cys Lys Arg Lys Leu Val Val Gly Asn Gly Pro Glu Gly Pro Gly  
     275                 280                285

Leu	Ser	Ser	Thr	Gly	Lys	Arg	Arg	Ala	Thr	Gly	Gln	Asp	Ser	Cys
				290		295					300			
Gln	Glu	Arg	Gly	Leu	Gln	Glu	Ala	Val	Arg	Arg	Glu	His	Gln	Glu
				305		310					315			
Ala	Ser	Val	Pro	Lys	Gly	Arg	Ile	Val	Pro	Arg	Gly	Ile	Asp	Arg
				320		325					330			
Leu	Glu	Arg	Thr	Arg	Ser	Ser	Arg	Lys	Ser	Lys	His	Gln	Glu	Ala
				335		340					345			
Thr	Glu	Thr	Ser	Leu	Leu	His	Ser	His	Arg	Phe	Lys	Lys	Gly	Gln
				350		355					360			
Lys	Leu	Gly	Lys	Asp	Ser	Phe	Pro	Thr	Gln	Asp	Leu	Thr	Pro	Leu
				365		370					375			
Gln	Asn	Ala	Cys	Phe	Trp	Thr	Lys	Thr	Arg	Ala	Ser	Phe	Ser	Phe
				380		385					390			
His	Lys	Lys	Lys	Ile	Val	Thr	Asp	Val	Ser	Glu	Val	Cys	Ser	Ile
				395		400					405			
Tyr	Thr	Thr	Ala	Thr	Ser	Leu	Ser	Gly	Ser	Leu	Leu	Ser	Glu	Cys
				410		415					420			
Ser	Asn	Arg	Pro	Val	Met	Asn	Arg	Thr	Ser	Gly	Ala	Pro	Ser	Ser
				425		430					435			
Trp	His	Ser	Ser	Ser	Met	Tyr	Leu	Leu	Ser	Pro	Leu	Asn	Thr	Leu
				440		445					450			
Ser	Ile	Ser	Asn	Lys	Lys	Ala	Ser	Asp	Ala	Glu	Lys	Val	Tyr	Gly
				455		460					465			
Glu	Cys	Ser	Gln	Lys	Gly	Pro	Val	Pro	Phe	Ser	His	Cys	Leu	Pro
				470		475					480			
Thr	Glu	Lys	Leu	Gln	Arg	Cys	Glu	Lys	Ile	Gly	Glu	Gly	Val	Phe
				485		490					495			
Gly	Glu	Val	Phe	Gln	Thr	Ile	Ala	Asp	His	Thr	Pro	Val	Ala	Ile
				500		505					510			
Lys	Ile	Ile	Ala	Ile	Glu	Gly	Pro	Asp	Leu	Val	Asn	Gly	Ser	His
				515		520					525			
Gln	Lys	Thr	Phe	Glu	Glu	Ile	Leu	Pro	Glu	Ile	Ile	Ile	Ser	Lys
				530		535					540			
Glu	Leu	Ser	Leu	Leu	Ser	Gly	Glu	Val	Cys	Asn	Arg	Thr	Glu	Gly
				545		550					555			
Phe	Ile	Gly	Leu	Asn	Ser	Val	His	Cys	Val	Gln	Gly	Ser	Tyr	Pro
				560		565					570			
Pro	Leu	Leu	Leu	Lys	Ala	Trp	Asp	His	Tyr	Asn	Ser	Thr	Lys	Gly
				575		580					585			
Ser	Ala	Asn	Asp	Arg	Pro	Asp	Phe	Phe	Lys	Asp	Asp	Gln	Leu	Phe
				590		595					600			
Ile	Val	Leu	Glu	Phe	Glu	Phe	Gly	Gly	Ile	Asp	Leu	Glu	Gln	Met
				605		610					615			
Arg	Thr	Lys	Leu	Ser	Ser	Leu	Ala	Thr	Ala	Lys	Ser	Ile	Leu	His
				620		625					630			
Gln	Leu	Thr	Ala	Ser	Leu	Ala	Val	Ala	Glu	Ala	Ser	Leu	Arg	Phe
				635		640					645			
Glu	His	Arg	Asp	Leu	His	Trp	Gly	Asn	Val	Leu	Leu	Lys	Lys	Thr
				650		655					660			
Ser	Leu	Lys	Lys	Leu	His	Tyr	Thr	Leu	Asn	Gly	Lys	Ser	Ser	Thr
				665		670					675			
Ile	Pro	Ser	Cys	Gly	Leu	Gln	Val	Ser	Ile	Ile	Asp	Tyr	Thr	Leu
				680		685					690			
Ser	Arg	Leu	Glu	Arg	Asp	Gly	Ile	Val	Val	Phe	Cys	Asp	Val	Ser
				695		700					705			
Met	Asp	Glu	Asp	Leu	Phe	Thr	Gly	Asp	Gly	Asp	Tyr	Gln	Phe	Asp
				710		715					720			
Ile	Tyr	Arg	Leu	Met	Lys	Lys	Glu	Asn	Asn	Asn	Arg	Trp	Gly	Glu
				725		730					735			
Tyr	His	Pro	Tyr	Ser	Asn	Val	Leu	Trp	Leu	His	Tyr	Leu	Thr	Asp
				740		745					750			
Lys	Met	Leu	Lys	Gln	Met	Thr	Phe	Lys	Thr	Lys	Cys	Asn	Thr	Pro

	755		760		765
Ala Met Lys Gln Ile	Lys Arg Lys Ile	Gln	Glu Phe His Arg	Thr	
	770		775		780
Met Leu Asn Phe Ser	Ser Ala Thr Asp	Leu	Leu Cys Gln His	Ser	
	785		790		795
Leu Phe Lys					

<210> 7  
<211> 632  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 5850786CD1

<400> 7

Met Phe Gly Lys Lys Lys Lys Ile Glu Ile Ser Gly Pro Ser			
1	5	10	15
Asn Phe Glu His Arg Val His Thr Gly Phe Asp Pro Gln Glu Gln			
	20	25	30
Lys Phe Thr Gly Leu Pro Gln Gln Trp His Ser Leu Leu Ala Asp			
	35	40	45
Thr Ala Asn Arg Pro Lys Pro Met Val Asp Pro Ser Cys Ile Thr			
	50	55	60
Pro Ile Gln Leu Ala Pro Met Lys Thr Ile Val Arg Gly Asn Lys			
	65	70	75
Pro Cys Lys Glu Thr Ser Ile Asn Gly Leu Leu Glu Asp Phe Asp			
	80	85	90
Asn Ile Ser Val Thr Arg Ser Asn Ser Leu Arg Lys Glu Ser Pro			
	95	100	105
Pro Thr Pro Asp Gln Gly Ala Ser Ser His Gly Pro Gly His Ala			
	110	115	120
Glu Glu Asn Gly Phe Ile Thr Phe Ser Gln Tyr Ser Ser Glu Ser			
	125	130	135
Asp Thr Thr Ala Asp Tyr Thr Thr Glu Lys Tyr Arg Glu Lys Ser			
	140	145	150
Leu Tyr Gly Asp Asp Leu Asp Pro Tyr Tyr Arg Gly Ser His Ala			
	155	160	165
Ala Lys Gln Asn Gly His Val Met Lys Met Lys His Gly Glu Ala			
	170	175	180
Tyr Tyr Ser Glu Val Lys Pro Leu Lys Ser Asp Phe Ala Arg Phe			
	185	190	195
Ser Ala Asp Tyr His Ser His Leu Asp Ser Leu Ser Lys Pro Ser			
	200	205	210
Glu Tyr Ser Asp Leu Lys Trp Glu Tyr Gln Arg Ala Ser Ser Ser			
	215	220	225
Ser Pro Leu Asp Tyr Ser Phe Gln Phe Thr Pro Ser Arg Thr Ala			
	230	235	240
Gly Thr Ser Gly Cys Ser Lys Glu Ser Leu Ala Tyr Ser Glu Ser			
	245	250	255
Glu Trp Gly Pro Ser Leu Asp Asp Tyr Asp Arg Arg Pro Lys Ser			
	260	265	270
Ser Tyr Leu Asn Gln Thr Ser Pro Gln Pro Thr Met Arg Gln Arg			
	275	280	285
Ser Arg Ser Gly Ser Gly Leu Gln Glu Pro Met Met Pro Phe Gly			
	290	295	300
Ala Ser Ala Phe Lys Thr His Pro Gln Gly His Ser Tyr Asn Ser			
	305	310	315
Tyr Thr Tyr Pro Arg Leu Ser Glu Pro Thr Met Cys Ile Pro Lys			
	320	325	330
Val Asp Tyr Asp Arg Ala Gln Met Val Leu Ser Pro Pro Leu Ser			

	335	340	345
Gly Ser Asp Thr Tyr Pro Arg Gly Pro Ala Lys Leu Pro Gln Ser			
	350	355	360
Gln Ser Lys Ser Gly Tyr Ser Ser Ser Ser His Gln Tyr Pro Ser			
	365	370	375
Gly Tyr His Lys Ala Thr Leu Tyr His His Pro Ser Leu Gln Ser			
	380	385	390
Ser Ser Gln Tyr Ile Ser Thr Ala Ser Tyr Leu Ser Ser Leu Ser			
	395	400	405
Leu Ser Ser Ser Thr Tyr Pro Pro Pro Ser Trp Gly Ser Ser Ser			
	410	415	420
Asp Gln Gln Pro Ser Arg Val Ser His Glu Gln Phe Arg Ala Ala			
	425	430	435
Leu Gln Leu Val Val Ser Pro Gly Asp Pro Arg Glu Tyr Leu Ala			
	440	445	450
Asn Phe Ile Lys Ile Gly Glu Gly Ser Thr Gly Ile Val Cys Ile			
	455	460	465
Ala Thr Glu Lys His Thr Gly Lys Gln Val Ala Val Lys Lys Met			
	470	475	480
Asp Leu Arg Lys Gln Gln Arg Arg Glu Leu Leu Phe Asn Glu Val			
	485	490	495
Val Ile Met Arg Asp Tyr His His Asp Asn Val Val Asp Met Tyr			
	500	505	510
Ser Ser Tyr Leu Val Gly Asp Glu Leu Trp Val Val Met Glu Phe			
	515	520	525
Leu Glu Gly Gly Ala Leu Thr Asp Ile Val Thr His Thr Arg Met			
	530	535	540
Asn Glu Glu Gln Ile Ala Thr Val Cys Leu Ser Val Leu Arg Ala			
	545	550	555
Leu Ser Tyr Leu His Asn Gln Gly Val Ile His Arg Asp Ile Lys			
	560	565	570
Ser Asp Ser Ile Leu Leu Thr Ser Asp Gly Arg Val Ser Ser Val			
	575	580	585
Leu Arg Gly Phe Leu Asp Leu Met Leu Val Arg Glu Pro Ser Gln			
	590	595	600
Arg Ala Thr Ala Gln Glu Leu Leu Gly His Pro Phe Leu Lys Leu			
	605	610	615
Ala Gly Pro Pro Ser Cys Ile Val Pro Leu Met Arg Gln Tyr Arg			
	620	625	630
His His			

<210> 8  
<211> 350  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7182343CD1

<400> 8

Met Arg Arg Arg Val Ile Ala Arg Pro Val Gly Ser Ser Val Arg			
1	5	10	15
Leu Lys Cys Val Ala Ser Gly His Pro Arg Pro Asp Ile Thr Trp			
	20	25	30
Met Lys Asp Asp Gln Ala Leu Thr Arg Pro Glu Ala Ala Glu Pro			
	35	40	45
Arg Lys Lys Lys Trp Thr Leu Ser Leu Lys Asn Leu Arg Pro Glu			
	50	55	60
Asp Ser Gly Lys Tyr Thr Cys Arg Val Ser Asn Arg Ala Gly Ala			
	65	70	75
Ile Asn Ala Thr Tyr Lys Val Asp Val Ile Gln Arg Thr Arg Ser			

80	85	90
Lys Pro Val Leu Thr Gly Thr His Pro	Val Asn Thr Thr Val	Asp
95	100	105
Phe Gly Gly Thr Thr Ser Phe Gln Cys	Lys Val Arg Ser Asp	Val
110	115	120
Lys Pro Val Ile Gln Trp Leu Lys Arg	Val Glu Tyr Gly Ala	Glu
125	130	135
Gly Arg His Asn Ser Thr Ile Asp Val	Gly Gly Gln Lys Phe	Val
140	145	150
Val Leu Pro Thr Gly Asp Val Trp Ser	Arg Pro Asp Gly Ser	Tyr
155	160	165
Leu Asn Lys Leu Leu Ile Thr Arg Ala	Arg Gln Asp Asp Ala	Gly
170	175	180
Met Tyr Ile Cys Leu Gly Ala Asn Thr	Met Gly Tyr Ser Phe	Arg
185	190	195
Ser Ala Phe Leu Thr Val Leu Pro Asp	Pro Lys Pro Pro Gly	Pro
200	205	210
Pro Val Ala Ser Ser Ser Ala Thr	Ser Leu Pro Trp Pro	Val
215	220	225
Val Ile Gly Ile Pro Ala Gly Ala Val	Phe Ile Leu Gly Thr	Leu
230	235	240
Leu Leu Trp Leu Cys Gln Ala Gln Lys	Lys Pro Cys Thr Pro	Ala
245	250	255
Pro Ala Pro Pro Leu Pro Gly His Arg	Pro Pro Gly Thr Ala	Arg
260	265	270
Asp Arg Ser Gly Asp Lys Asp Leu Pro	Ser Leu Ala Ala Leu	Ser
275	280	285
Ala Gly Pro Gly Val Gly Leu Cys Glu	Glu His Gly Ser Pro	Ala
290	295	300
Ala Pro Gln His Leu Leu Gly Pro Gly	Pro Val Ala Gly Pro	Lys
305	310	315
Leu Tyr Pro Lys Leu Tyr Thr Asp Ile	His Thr His Thr His	Thr
320	325	330
His Ser His Thr His Ser His Val Glu	Gly Lys Val His Gln	His
335	340	345
Ile His Tyr Gln Cys		
350		

<210> 9  
<211> 571  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7264345CD1

<400> 9  
Met Asp Arg Ser Lys Arg Asn Ser Ile Ala Gly Phe Pro Pro Arg  
1 5 10 15  
Val Glu Arg Leu Glu Glu Phe Glu Gly Gly Gly Gly Glu Gly  
20 25 30  
Asn Val Ser Gln Val Gly Arg Val Trp Pro Ser Ser Tyr Arg Ala  
35 40 45  
Leu Ile Ser Ala Phe Ser Arg Leu Thr Arg Leu Asp Asp Phe Thr  
50 55 60  
Cys Glu Lys Ile Gly Ser Gly Phe Phe Ser Glu Val Phe Lys Val  
65 70 75  
Arg His Arg Ala Ser Gly Gln Val Met Ala Leu Lys Met Asn Thr  
80 85 90  
Leu Ser Ser Asn Arg Ala Asn Met Leu Lys Glu Val Gln Leu Met  
95 100 105  
Asn Arg Leu Ser His Pro Asn Ile Leu Arg Phe Met Gly Val Cys

	110		115		120
Val His Gln Gly	Gln	Leu His Ala Leu	Thr Glu Tyr Ile Asn	Ser	
	125		130		135
Gly Asn Leu Glu	Gln	Leu Leu Asp Ser	Asn Leu His Leu Pro	Trp	
	140		145		150
Thr Val Arg Val	Lys	Leu Ala Tyr Asp	Ile Ala Val Gly Leu	Ser	
	155		160		165
Tyr Leu His Phe	Lys	Gly Ile Phe His	Arg Asp Leu Thr Ser	Lys	
	170		175		180
Asn Cys Leu Ile	Lys	Arg Asp Glu Asn	Gly Tyr Ser Ala Val	Val	
	185		190		195
Ala Asp Phe Gly	Leu	Ala Glu Lys Ile	Pro Asp Val Ser Met	Gly	
	200		205		210
Ser Glu Lys Leu	Ala	Val Val Gly Ser	Pro Phe Trp Met Ala	Pro	
	215		220		225
Glu Val Leu Arg	Asp	Glu Pro Tyr Asn	Glu Lys Ala Asp Val	Phe	
	230		235		240
Ser Tyr Gly Ile	Ile	Leu Cys Glu Ile	Ile Ala Arg Ile Gln	Ala	
	245		250		255
Asp Pro Asp Tyr	Leu	Pro Arg Thr Glu	Asn Phe Gly Leu Asp	Tyr	
	260		265		270
Asp Ala Phe Gln	His	Met Val Gly Asp	Cys Pro Pro Asp Phe	Leu	
	275		280		285
Gln Leu Thr Phe	Asn	Cys Cys Asn Met	Asp Pro Lys Leu Arg	Pro	
	290		295		300
Ser Phe Val Glu	Ile	Gly Lys Thr Leu	Glu Glu Ile Leu Ser	Arg	
	305		310		315
Leu Gln Glu Glu	Glu	Gln Glu Arg Asp	Arg Lys Leu Gln Pro	Thr	
	320		325		330
Ala Arg Gly Leu	Leu	Glu Lys Ala Pro	Gly Val Lys Arg Leu	Ser	
	335		340		345
Ser Leu Asp Asp	Lys	Ile Pro His Lys	Ser Pro Cys Pro Arg	Arg	
	350		355		360
Thr Ile Trp Leu	Ser	Arg Ser Gln Ser	Asp Ile Phe Ser Arg	Lys	
	365		370		375
Pro Pro Arg Thr	Val	Ser Val Leu Asp	Pro Tyr Tyr Arg Pro	Arg	
	380		385		390
Asp Gly Ala Ala	Arg	Thr Pro Lys Val	Asn Pro Phe Ser Ala	Arg	
	395		400		405
Gln Asp Leu Met	Gly	Gly Lys Ile Lys	Phe Phe Asp Leu Pro	Ser	
	410		415		420
Lys Ser Val Ile	Ser	Leu Val Phe Asp	Leu Asp Ala Pro Gly	Pro	
	425		430		435
Gly Thr Met Pro	Leu	Ala Asp Trp Gln	Glu Pro Leu Ala Pro	Pro	
	440		445		450
Ile Arg Arg Trp	Arg	Ser Leu Pro Gly	Ser Pro Glu Phe Leu	His	
	455		460		465
Gln Glu Ala Cys	Pro	Phe Val Gly Arg	Glu Glu Ser Leu Ser	Asp	
	470		475		480
Gly Pro Pro Pro	Arg	Leu Ser Ser Leu	Lys Tyr Arg Val Lys	Glu	
	485		490		495
Ile Pro Pro Phe	Arg	Ala Ser Ala Leu	Pro Ala Ala Gln Ala	His	
	500		505		510
Glu Ala Met Asp	Cys	Ser Ile Leu Gln	Glu Glu Asn Gly Phe	Gly	
	515		520		525
Ser Arg Pro Gln	Gly	Thr Ser Pro Cys	Pro Ala Gly Ala Ser	Glu	
	530		535		540
Glu Met Glu Val	Glu	Glu Arg Pro Ala	Gly Ser Thr Pro Ala	Thr	
	545		550		555
Phe Ser Thr Ser	Gly	Ile Gly Leu Gln	Thr Gln Gly Lys Gln	Asp	
	560		565		570
Gly					

<210> 10  
<211> 323  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 7472328CD1

<400> 10

Met	Arg	Ala	Thr	Pro	Leu	Ala	Ala	Ser	Ala	Gly	Ser	Leu	Ser	Arg
1					5					10				15
Lys	Lys	Arg	Leu	Glu	Leu	Asp	Asp	Asn	Leu	Asp	Thr	Glu	Arg	Pro
					20					25				30
Val	Gln	Lys	Arg	Ala	Arg	Ser	Gly	Pro	Gln	Pro	Arg	Leu	Pro	Pro
					35					40				45
Cys	Leu	Leu	Pro	Leu	Ser	Pro	Pro	Thr	Ala	Pro	Asp	Arg	Ala	Thr
					50					55				60
Ala	Val	Ala	Thr	Ala	Ser	Arg	Leu	Gly	Pro	Tyr	Val	Leu	Leu	Glu
					65					70				75
Pro	Glu	Glu	Gly	Gly	Arg	Ala	Tyr	Gln	Ala	Leu	His	Cys	Pro	Thr
					80					85				90
Gly	Thr	Glu	Tyr	Thr	Cys	Lys	Val	Tyr	Pro	Val	Gln	Glu	Ala	Leu
					95					100				105
Ala	Val	Leu	Glu	Pro	Tyr	Ala	Arg	Leu	Pro	Pro	Asp	Lys	His	Val
					110					115				120
Ala	Trp	Pro	Thr	Glu	Gly	Leu	Ala	Gly	Thr	Gln	Leu	Leu	Tyr	Ala
					125					130				135
Phe	Phe	Thr	Trp	Thr	His	Gly	Asp	Met	His	Arg	Leu	Ile	Gly	His
					140					145				150
Thr	Pro	Cys	Ala	His	Cys	Asp	Gln	Thr	Arg	Lys	Lys	Leu	Val	Leu
					155					160				165
Glu	Asn	Leu	Glu	Asp	Ser	Cys	Val	Leu	Thr	Gly	Pro	Asp	Asp	Ser
					170					175				180
Leu	Trp	Asp	Lys	His	Ala	Cys	Pro	Ala	Tyr	Val	Gly	Pro	Glu	Ile
					185					190				195
Leu	Ser	Ser	Arg	Ala	Ser	Tyr	Ser	Gly	Lys	Ala	Ala	Asn	Val	Trp
					200					205				210
Ser	Leu	Gly	Val	Ala	Leu	Phe	Thr	Met	Leu	Ala	Gly	His	Tyr	Pro
					215					220				225
Phe	Gln	Asp	Ser	Glu	Pro	Val	Leu	Leu	Phe	Gly	Lys	Ile	Arg	Arg
					230					235				240
Gly	Ala	Tyr	Ala	Leu	Pro	Ala	Gly	Leu	Ser	Ala	Pro	Ala	Arg	Cys
					245					250				255
Leu	Val	Arg	Cys	Leu	Leu	Arg	Arg	Glu	Pro	Ala	Glu	Arg	Leu	Thr
					260					265				270
Ala	Thr	Gly	Ile	Leu	Leu	His	Pro	Trp	Leu	Arg	Gln	Asp	Pro	Met
					275					280				285
Pro	Leu	Ala	Pro	Thr	Arg	Ser	His	Leu	Trp	Glu	Ala	Ala	Gln	Val
					290					295				300
Val	Pro	Asp	Gly	Leu	Gly	Leu	Asp	Glu	Ala	Arg	Glu	Glu	Gly	
					305					310				315
Asp	Arg	Glu	Val	Val	Leu	Tyr	Gly							
					320									

<210> 11  
<211> 471  
<212> PRT  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 1471412CD1

&lt;400&gt; 11

Met	Glu	Asn	Leu	Lys	His	Ile	Ile	Thr	Leu	Gly	Gln	Val	Ile	His
1				5					10				15	
Lys	Arg	Cys	Glu	Glu	Met	Lys	Tyr	Cys	Lys	Lys	Gln	Cys	Arg	Arg
				20					25				30	
Leu	Gly	His	Arg	Val	Leu	Gly	Leu	Ile	Lys	Pro	Leu	Glu	Met	Leu
				35					40				45	
Gln	Asp	Gln	Gly	Lys	Arg	Ser	Val	Pro	Ser	Glu	Lys	Leu	Thr	Thr
				50					55				60	
Ala	Met	Asn	Arg	Phe	Lys	Ala	Ala	Leu	Glu	Glu	Ala	Asn	Gly	Glu
				65					70				75	
Ile	Glu	Lys	Phe	Ser	Asn	Arg	Ser	Asn	Ile	Cys	Arg	Phe	Leu	Thr
				80					85				90	
Ala	Ser	Gln	Asp	Lys	Ile	Leu	Phe	Lys	Asp	Val	Asn	Arg	Lys	Leu
				95					100				105	
Ser	Asp	Val	Trp	Lys	Glu	Leu	Ser	Leu	Leu	Leu	Gln	Val	Glu	Gln
				110					115				120	
Arg	Met	Pro	Val	Ser	Pro	Ile	Ser	Gln	Gly	Ala	Ser	Trp	Ala	Gln
				125					130				135	
Glu	Asp	Gln	Gln	Asp	Ala	Asp	Glu	Asp	Arg	Arg	Ala	Phe	Gln	Met
				140					145				150	
Leu	Arg	Arg	Asp	Asn	Glu	Lys	Ile	Glu	Ala	Ser	Leu	Arg	Arg	Leu
				155					160				165	
Glu	Ile	Asn	Met	Lys	Glu	Ile	Lys	Glu	Thr	Leu	Arg	Gln	Tyr	Leu
				170					175				180	
Pro	Pro	Lys	Cys	Met	Gln	Glu	Ile	Pro	Gln	Glu	Gln	Ile	Lys	Glu
				185					190				195	
Ile	Lys	Lys	Glu	Gln	Leu	Ser	Gly	Ser	Pro	Trp	Ile	Leu	Leu	Arg
				200					205				210	
Glu	Asn	Glu	Val	Ser	Thr	Leu	Tyr	Lys	Gly	Glu	Tyr	His	Arg	Ala
				215					220				225	
Pro	Val	Ala	Ile	Lys	Val	Phe	Lys	Lys	Leu	Gln	Ala	Gly	Ser	Ile
				230					235				240	
Ala	Ile	Val	Arg	Gln	Thr	Phe	Asn	Lys	Glu	Ile	Lys	Thr	Met	Lys
				245					250				255	
Lys	Phe	Glu	Ser	Pro	Asn	Ile	Leu	Arg	Ile	Phe	Gly	Ile	Cys	Ile
				260					265				270	
Asp	Glu	Thr	Val	Thr	Pro	Pro	Gln	Phe	Ser	Ile	Val	Met	Glu	Tyr
				275					280				285	
Cys	Glu	Leu	Gly	Thr	Leu	Arg	Glu	Leu	Leu	Asp	Arg	Glu	Lys	Asp
				290					295				300	
Leu	Thr	Leu	Gly	Lys	Arg	Met	Val	Leu	Val	Leu	Gly	Ala	Ala	Arg
				305					310				315	
Gly	Leu	Tyr	Arg	Leu	His	His	Ser	Glu	Ala	Pro	Glu	Leu	His	Gly
				320					325				330	
Lys	Ile	Arg	Ser	Ser	Asn	Phe	Leu	Val	Thr	Gln	Gly	Tyr	Gln	Val
				335					340				345	
Lys	Leu	Ala	Gly	Phe	Glu	Leu	Arg	Lys	Thr	Gln	Thr	Ser	Met	Ser
				350					355				360	
Leu	Gly	Thr	Thr	Arg	Glu	Lys	Thr	Asp	Arg	Val	Lys	Ser	Thr	Ala
				365					370				375	
Tyr	Leu	Ser	Pro	Gln	Glu	Leu	Glu	Asp	Val	Phe	Tyr	Gln	Tyr	Asp
				380					385				390	
Val	Lys	Ser	Glu	Ile	Tyr	Ser	Phe	Gly	Ile	Val	Leu	Trp	Glu	Ile
				395					400				405	
Ala	Thr	Gly	Asp	Ile	Pro	Phe	Gln	Gly	Cys	Asn	Ser	Glu	Lys	Ile
				410					415				420	
Arg	Lys	Leu	Val	Ala	Val	Lys	Arg	Gln	Gln	Glu	Pro	Leu	Gly	Glu
				425					430				435	
Asp	Cys	Pro	Ser	Glu	Leu	Arg	Glu	Ile	Ile	Asp	Glu	Cys	Arg	Ala
				440					445				450	
His	Asp	Pro	Ser	Val	Arg	Pro	Ser	Val	Asp	Glu	Ile	Leu	Lys	Lys
				455					460				465	

Leu Ser Thr Phe Ser Lys  
470

<210> 12  
<211> 4335  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 4912053CB1

<400> 12

atgaaaagagc	acttcttcag	tgcctttc	agtatacaa	agtaaaaacc	aggtaactgtt	60
tctgaaggat	tgtcattctt	gcatagcagt	gtgaaaatgg	tgcataaaa	tatcaactcct	120
gaaaatataa	tttgaataa	aagtggagcc	tggaaaataa	tgggtttga	ttttgtgt	180
tcatcaacca	atccttctga	acaagagcct	aaatttcctt	gtaaaagaatg	ggacccaaat	240
ttaccttcat	tgtgtcttcc	aaatcctgaa	tatgggctc	ctgaatacat	actttctgtg	300
agctgtaaa	cagccagtga	tatgtattct	tttaggaactg	ttatgtatgc	tgtatttaat	360
aaaggaaac	ctatatttga	agtcaacaag	caagatattt	acaagagttt	cagtaggcag	420
ttggatcagt	tgagtcgtt	aggatctgt	tcacttacaa	atatacctga	ggaagttcg	480
gaacatgtaa	agctactgtt	aatgttaact	ccgactgtaa	gaccagatgc	agatcaaatg	540
acaaagattc	ccttcttga	tgtgttggt	gcagtaacac	tgcaatattt	tgataccta	600
ttccaaagag	ataatctca	gaaatcacag	tttttcaaag	gactgccaaa	ggttctacca	660
aaactgccc	agcgtgtcat	tgtcagaga	atttgcctt	gtttgactt	agaatttgta	720
aaccctgaca	tggtacctt	tgtttgccc	aatgtctac	ttattgctga	ggaatgcacc	780
aaagaagaat	atgtcaaatt	aattcttcct	gaacttggcc	ctgtgtttaa	gcagcaggag	840
ccaatccagg	ctagcaacat	gattttgtt	atttcctac	aaaaaaatgga	tttgcata	900
accaaaaaccc	ctcctgatga	gataaagaac	agtgttctac	ccatggttt	cagagcacta	960
gaagctcctt	ccattcagat	ccaggagctc	tgtctaaaca	tcattccaaac	ctttgcaat	1020
cttatactact	acccatccat	gaaaaacgct	ttgataccaa	gaattaaaaaa	tgcttgtcta	1080
caaacatctt	cccttgcgt	tcgtgtaaat	tcatttagt	gcttaggaaa	gattttggaa	1140
tacttggata	agtggtttgt	acttgcgtat	atcctaccct	tcttacaaca	aattccatcc	1200
aaggaacctg	cggtcctcat	gggaatttta	ggtatttaca	aatgtacttt	tactcataag	1260
aagttggaa	tcaccaaaga	gcagctggcc	ggaaaagtgt	tgccctcatct	tattccctg	1320
agtattgaaa	acaatcttaa	tcttaatcag	ttcaattctt	tcatttccgt	cataaaagaa	1380
atgcttaata	gattggagtc	tgaacataag	actaaactgg	agcaacttca	tataatgcaa	1440
gaacagcaga	aatctttgga	tataggaaat	caaataatg	tttctgagga	gatgaaagtt	1500
acaaatattg	ggaatcagca	aattgacaaa	tttttaaca	acattggagc	agaccttctg	1560
actggcagtg	agtccgaaaa	taaagaggac	gggttacaga	ataaacataa	aagagcatca	1620
cttacacttg	aagaaaaaca	aaaattagca	aaagaacaag	agcaggcaca	gaagctgaaa	1680
agccagcagc	ctcttaaacc	ccaagtgcac	acacctgtt	ctactgtt	acagactaag	1740
gacttgcac	acacactgt	ggataatatg	tcatccttga	ccagccttc	tgttagtacc	1800
cctaaatctt	ctgcttcaag	tactttact	tctgttccctt	ccatggcat	tggtatgt	1860
tttctacac	caactgataa	tacaaagaga	aatttgcacaa	atggcctaaa	tgccaatatg	1920
ggctttcaga	cttcaggatt	caacatgccc	gttaatacaa	accagaactt	ctacagtatg	1980
ccaagcacag	ttggagtgac	caagatgact	ctggaaacac	ctcccactt	gccaaacttc	2040
aatgcttga	gtgttcctcc	tgctggtgca	aagcagaccc	aacaaagacc	cacagatatg	2100
tctgccctt	ataatctctt	tggccctcag	aaacccaaag	ttagcatgaa	ccagttatca	2160
caacagaaac	caaatcagt	gcttaatcag	tttgcgtatc	ctcaaggttc	tccaaactatg	2220
ggcagttcag	taatggggac	acagatgaa	gtgataggac	aatctgctt	tggtatgcag	2280
ggttaatcctt	tcttaaacc	acagaacttt	gcacagccac	caactactat	gaccaatagc	2340
agttcagcta	gcaatgattt	aaaagatctt	tttgggtgag	gtgtcttact	tctattttga	2400
aggattattt	cagttcaat	catgggtgag	ctgatttaca	tctttatata	gttggcttgg	2460
aggaagtact	tctatggaa	agtgaacagt	tctgtgacag	gaaacatctc	tgtccatgcc	2520
agcatagtag	ttgttatggac	ttctaaccag	ttgagttttt	taaagcattt	aggatttttt	2580
cctcttacca	actccttttc	aggttttaa	agacccagcc	cttcccaatc	tcaaagagaa	2640
aaaggaaact	gagttatctt	gaataacata	acttttaat	caaatgttta	tttggcttgg	2700
tggatcttgg	tgttattttaa	aaaattgagg	tgatggtcat	tgcaagctca	tctattaagt	2760
actatatgtt	acacagtcta	tgagtcatta	gtcttcattt	taatatgtaa	aaaatcttga	2820
tgctgtattt	atttgcgttgc	atttaagatg	acagtggagaa	aatgataagc	ataaagagaa	2880
gtatcaggtt	atttgccttt	tccaaactt	tcagatgaac	tattgtttag	tacagagact	2940
gagcaaatac	tacaaaattc	aacttaaccc	tcatttcattt	ggtttaatg	cgttatataac	3000

catcttaagt gcaaactaat cattgtaaat tatattttag catggctgc ctcaaatag 3060  
aatgtatccc tctgcattca cttggatata tttagaatca ctttttcct cctgtatcaa 3120  
ggaagaggta tgtgctgatt tgtttggata tttgacaagg cactctgatg tgacttccct 3180  
gactactacc ttcatatttc atttcaaatt caaacttctg aggttgcagc atatatgaat 3240  
tgcatatcca aaagaagatt tgtaagaatt aaactatatt tatgagtaaa cttttgaggt 3300  
ttctgctgta ttgtttcaaa tgtaataaac ttacttctg taaaaattga gcagttgtat 3360  
cttctgacca ccaacagatt ttcagcttgc catgatagtc tgacccatt aattactgct 3420  
actgaagttc aatttttttc taggaatttt aggaacctt tgtttaataa ttttaatttc 3480  
tattagccat ttttaggaag gaaagaatca attctcttaa cagaaaacat gctttatccc 3540  
tcaaaacctt tctctgatat tttcttttaa tttgctgatt attcaaccac aggccttat 3600  
gctataaatg tcatttgtat tttaaaaaat aatattccac tcataaaact taaaaccat 3660  
cttcaacga actatatacg tattatagtt gctgccatag agttgatggt ttttaattat 3720  
ctggaaccag caatcattta aaataaaacca tattaagtt agtatgctgg tattgtttat 3780  
tcattttata tgaatattca ttgaaaatat atacacaata tatgtaatac acagcacttg 3840  
attacaaaaat gtaatttaat tatattattg ctggcagcat tcagttaga gggtaacttta 3900  
aaaaatagaa gtcagcttc acatctgatt tctgtatggg ctgtacttgg ttaacttgat 3960  
tttagaaaaaa ggactaacag aattgctaaa gaaatgcac caataaaatga aaaacagtag 4020  
gaagatcaaa tgaaaaatgc aaatataattc acaacttgc cagattagct gtcctgtttg 4080  
taatgcaata ttaatatgtc ttttggaaa aagcctacat atggaataaa ataagtattg 4140  
aagaattttt ctttgtaaca atttagtagt cactgtttat tgagaaaattg ttttttattt 4200  
tgtaaaataa catgatgtaa gtgttgaccc cttaaacaga gagagcgctt aatatcaccg 4260  
catatagctc ttgacccctt atatgtgaat atgacacatg tgcaaaaatac agtgtgtgaa 4320  
atctgtgcac acgct 4335

<210> 13  
<211> 1829  
<212> DNA  
<213> Homo

<220>  
<221> misc\_feature  
<223> Incyte ID No: 870615CB1

<400> 13  
gcggctgcat ctgaggcgga cgccgcagtg tagggcacac cctgagggta gcgagagacg 60  
tgacaggacc gagcagctgg gggctgcagc ctgctctgtc cccaccgatc ccagagagag 120  
gctgtgagtc tggatctggg cgttcgacc ctgagagagg gggtcaaagc tcccctgcgg 180  
ccccgaggaa gaagagcaag gttgaggaca gcctcggtc accagagtag tttggaggca 240  
ggcctgcagg cgctccccg ctcctcagct ccaagctgac gtatctacac ctgccggcgg 300  
gcattaagat ggctggttat gccactactc ccagccccat gcagaccctt caggaggaag 360  
cggtgtgtgc catctgcttg gattacttca aggaccccgt gtccatcagc tgtgggcaca 420  
acttctgccg agggtgtgtg acccagctgt ggagtaagga ggacgaggag gcccagaacg 480  
aggaggaaga tgaatggag gaggaggagg acgaggaagc ggtggggca tggatggatg 540  
ggacggctcc attcgagagg tggatgtatcg gggaatgct gacgaagagt tgttccaaga 600  
ccaagatgac gatgaactct ggctcggtga cagtggata actaattggg acaacgtaga 660  
ctatatgtgg gacgaggagg aagaagaaga ggaagatcag gactattacc taggaggctt 720  
gagacctgac ctgagaattg atgtctaccg agaagaagaa atactggaag catacgatga 780  
ggacgaagat gaagagctgt atcctgacat ccacccgcct cttccttgc cccttccagg 840  
gcagttcacc tgcccccaagt gccgaaagag ctttacacgt cgtagcttc gtcccaactt 900  
gcagctggcc aacatggtcc agataattcg ccagatgtgc cccactcctt atcggggaaa 960  
ccggagtaat gatcaggca tgtgtttaa acaccagga gccctgaaac tcttctgtga 1020  
ggtggacaaa gaggccatct gtgtgggtgt ccgagaatcc aggagccaca aacagcacag 1080  
cgtggtgcca ttggaggagg tggtgcagga gtaccagaag atagggtcaa cttcttcagt 1140  
tgagttatct gaaagtgttgc ggcagtgaca ccatgccaaa cactgttaaa ttcatgccat 1200  
taaaaacagt ggagtgttct gcaaagcccg attctctgca gctttaagac tggacagtat 1260  
tgaaatattc acaggaatct tccaaagccgt gaaagcttaa tattaaacag cccttttaat 1320  
tgcaaaggag aaaaaaatgg agacacttgt gaaaccttgc attctgagtg ctgccacaac 1380  
taaattaagg aattccagaa tttcttcatc tacttctgca gtcttaactc tgaattaatg 1440  
actgttatata tcaaacttca gatcttaaaa aacatcagct cgcaagaggg tgaggcagga 1500  
gaatgacatg aacctggag gcggagcttgc cagtaagccg agactgagcc actgcactcc 1560  
agcctggca atagagcaag actccatctc aaaaaaaaaa aaaaatcggc tcattttttta 1620  
ttgctgttgg ttgagatgaa tctactgtt aagaaaaactg tactttgctt acattcagaa 1680  
taaggcttca aaaactgtgg gtgtgctgtt ttctgtgagcc ttctacttga ccaccaaaga 1740

cagtatttt gcagctgtat caaatgaacc gagttggagg aaattgtata ggagtatatc 1800  
 ttaatggcac tacagcaaac tgtgttaat 1829

<210> 14  
 <211> 2143  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 2762870CB1

<400> 14  
 ggacaacatg gctgcgccccg cactaggct ggtgtgtgga cgttgccctg agctgggtct 60  
 cgtcctcttg ctgctgctgc tctcgctgct gtgtggagcg gcagggagcc aggaggccgg 120  
 gaccgggtgcg ggcgcggggt cccttgcggg ttcttgcggc tgccgcacgc cccagcggcc 180  
 tggcgcccat ggcagttcgg cagccgctca ccgataactcg cgggaggcta acgctccggg 240  
 ccccgtaccc ggagagcggc aactcgcgca ctcaaagatg gtcccccattcc ctgctggagt 300  
 atttacaatg ggcacagatg atcctcagat aaagcaggat gggaaagcac ctgcgaggag 360  
 agttactatt gatgccttt acatggatgc ctatgaagtc agtaatactg aatttgagaa 420  
 gtttgtgaac tcaactggct atttgacaga ggctgagaag tttggcact ctttgtctt 480  
 tgaaggcatg ttgagtgagc aagtgaagac caatattcaa caggcagttg cagctgctcc 540  
 ctgggtggta cctgtgaaaag ggcgctaactg gagacaccca gaaggccctg actctactat 600  
 tctgcacagg ccggatcatc cagttctcca tgtgtcctgg aatgatgcgg ttgcctactg 660  
 cacttggcca gggaaagcggc tgcccacgga agctgagtgg gaatacagct gtcgaggagg 720  
 cctgcataat agactttcc cctggggcaa caaactgcag cccaaaggcc agcattatgc 780  
 caacatttgg cagggcgagt ttccggtgac caacactggt gaggatggct tccaaggaac 840  
 tgcgcctgtt gatgccttcc ctcccaatgg ttatggctt tacaacatag tggggAACGC 900  
 atgggaatgg acttcagact ggtggactgt tcatttattt gttgaagaaa cgcttaacc 960  
 aaaaggtccc cttctggga aagaccgagt gaagaaaggt ggatcctaca tgtgccatag 1020  
 gtcttattgt tacaggtatc gctgtgtgc tcggagccag aacacacctg atagctctgc 1080  
 ttcaaatctg ggattccgct gtgcagccga ccgcctgccc actatggact gacaaccaag 1140  
 gaaagtcttc cccagtccaa ggagcagtcg tgtctgacct acattggct tttctcagaa 1200  
 ctttgaacga tcccatgcaa agaattccca ccctgaggtg gtttacatac ctgcccata 1260  
 gccaaaggaa ccgccttgg agaccaaatt gctgacctgg gtcagtgcatt gtgcattatg 1320  
 gtgtggtgca tctttggaga tcatcgccat attttacttt tgagagtctt taaagaggaa 1380  
 ggggagtggaa gggaaacctg agctaggctt caggaggccc gcgtcctacg caggctctgc 1440  
 cacaggggtt agaccccagg tccgacgctt gacccctctg ggcctcaagt gccctccct 1500  
 atcaaataatgaa gggatggaca gcatgacctc tgggtgtctc tccaaactcac cagtctaaa 1560  
 aagggtatca gattctattt tgacttcata gtgagaattt atgatagatt atttttagc 1620  
 tatttttcc atgtgtgaac cttgagtgat actaatcatg taaagtaaga gttctcttat 1680  
 gtattatttt cgaaagaggg gtgtggtgac tcctttatat tcgtactgca ctttgggttt 1740  
 ccaaggaaat cagtgtctt tacgttggta tcatgtatcc cacatggggc cggtgatgg 1800  
 atgctgcagt tcagccgttg aacacatagg aatgtctgtg gggtgactct actgtgcttt 1860  
 atcttttaac attaagtgcc tttgggttcag aggggcagtc ataagctctg tttccccctc 1920  
 tccccaaagc cttcagcgaa cgtgaaatgt gcgctaaacg gggaaacctg tttaattcta 1980  
 gatataaggaa aaaaggaacg aggaccttga atgagctata ttcagggtat ccggtatttt 2040  
 gtaataggaa ataggaaacc ttgttggctg tggaaatatcc gatgcttga atcatgcact 2100  
 gtgttgaata aacgtatctg ctaaaatcagg ctccaccgtc aga 2143

<210> 15  
 <211> 2310  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 5507012CB1

<400> 15  
 ccgagtgacc gggagccggg ccgcggcctt ccctcgcccc cctcgcccccc tcccactcct 60  
 ctgccccggg gccgccaccgc cccgggcgtc ggacctggtc ccgtgctcgc ggtgccgccc 120  
 ccctctgggc ctagcccgcc cagctcgccg agcggccggca gtgggagccg cgtccgccc 180

atccgcctcg actcggtgcc ggccccctggc cctccctca tgactgcggc gcctctgctg 240  
 ccaccgcccc cccggccgccc gctcgccgca ggatggatgc ggaccgtgcg gcgctaacc 300  
 ccgtggctca gctcccgaaat cggccgcctt cgagccctcc tcgtgagccg cagcagcctc 360  
 ggtgccagcc cccgcccag ctggggccag cggtcgcct gtccctcggt gcggcttg 420  
 ggtgctgagt gaggcgtcgt cgggtcggc gcgaacccgc cggccgcgg ttccctgcag 480  
 acctctgcgc gggcggtcg gcccTTcacg ccctttcggt tcacgaatcc gagccgc 540  
 gcctctctcc agcgaaccga ccatgtctgg cggccgcga gagaagcaga gcagcactcc 600  
 cggttccctg ttccctctcg cggccgcctc tgcccccaag aatggctcca gctccgattc 660  
 ctccgtgggg gagaaaactgg gagccgcggc cgccgacgct gtgaccggca ggaccgagga 720  
 gtacaggcgc cgccgcacaca ctatggacaa ggacagccgt gggccggccg cgaccactac 780  
 caccactgag caccgcttct tccgcccggag cgtcatctgc gactccaatg ccactgcact 840  
 ggagcttccc ggccttctc ttccctgccc ccagcccagc atccccgcgg ctgtcccgca 900  
 gagtgctcca cccggagcccc accgggaaga gaccgtgacc gccaccgc 960  
 agcccagcag cctccagccg ctgcccggcc tggggAACAG gccgtcgccg gccctgc 1020  
 ctcgactgtc cccagcagta ccagcaaaga ccgcccagtgc tcccagccta gccttgtgg 1080  
 gagcaaagag gagccgcgc cggcgagaag tggcagcggc ggcggcagcg ccaaggagcc 1140  
 acaggaggaa cggagccagc agcaggatga tatcgaagag ctggagacca aggccgtggg 1200  
 aatgtctaac gatggccgct ttctcaagtt tgacatcgaa atcggcagag gctccttta 1260  
 gacggtctac aaaggtctgg acactgaaac caccgtggaa gtcgcctgggt gtgaactgca 1320  
 ggatcgaaaa ttaacaaagt ctgagaggca gagattaaa gaagaagctg aaatgttaaa 1380  
 aggtcttcag catcccaata ttgttagatt ttatgattcc tgggaatcca cagtaaaagg 1440  
 aaagaagtgc attgtttgg tgactgaact tatgacgtct ggaacactta aaacgttatc 1500  
 gaaaagggtt aaagtgtatga agatcaaagt tctaagaagc tggtgccgtc agatcctta 1560  
 aggtcttcag tttcttcata ctcgaactcc acctatcatt caccgcgtc ttaaatgtga 1620  
 caacatctt atcaccggcc ctactggctc agtcaagatt ggagacctcg gtctggcaac 1680  
 cctgaagcgg gcttctttg ccaagagtgt gataggtaacc ccagagttca tggccctga 1740  
 gatgtatgag gagaatatg atgaatccgt tgacgtttat gctttggga tgtgcattgc 1800  
 tgagatggct acatctgaat atccttactc ggagtgc 1860  
 tcgcgtgacc agtgggtga agccagccag ttttgc 1920  
 ggaaattatt gaaggatgca tacgacaaaa caaagatgaa agatattcca tcaaagac 1980  
 tttgaaccat gccttcttcc aagagggaaac aggatgtacgg gtagaattag cagaagaaga 2040  
 ttagatggagaa aaaatagcca taaaattatg gctacgtatt gaagatatta agaaattaaa 2100  
 gggaaaatac aaagataatg aagctattga gttttttt gatttagaga gagatgtccc 2160  
 agaagatgtt gcacaagaaaa tggtagagtc tgggtatgtc tgtgaagggtg atcacaagac 2220  
 catggctaaa gcttcaaag acagagtatc attaatttag gggaaacgag agcagcggca 2280  
 atttgtacgg gaggagccag gaaaaaaaaa 2310

<210> 16  
 <211> 4699  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 6756529CB1

<400> 16

gcagcagcgg cggcgccagc cggagcagta ggcacccgag cagcgccagc ggccgagcgg 60  
 gcggtttcct ggcctggcg ctccgggtggc ggccggaggtg cgccgcggagc catgttata 120  
 atgtcggagt tcagcgcgga ccccgccggc cagggtcagg gccagcagaa gcccctccgg 180  
 gtgggtttt acgacatcga gcggaccctg ggcaaaaggca acttcgcgtt ggtgaagctg 240  
 ggcggccatc gagtcaccaa aacgcaggtt gcaataaaaa taattgataa aacacgatta 300  
 gattcaagca atttggagaa aatctatcgt gaggttcagc tgatgaagct tctgaaccat 360  
 ccacacatca taaagcttta ccaggttatg gaaacaaagg acatgctta catcgtaact 420  
 gaatttgcta aaaatggaga aatgtttgat tatttgactt ccaacggca cctgagttag 480  
 aacgaggcgc ggaagaagtt ctggcaaattc ctgtcgccg tggagtactg tcacgaccat 540  
 cacatcgccc accgggacct caagaccgag aacctcctgc tggatggcaa catggacatc 600  
 aagctggcag attttggatt tggaaatttc tacaagtcag gagacgtct gtccacgtgg 660  
 tgtgggagcc ccccgatgc cgccccggaa gtcttgagg ggaaggagta tgaaggcccc 720  
 cagctggaca tctggagcct gggcgtgggtg ctgtacgtcc tggatggcc ttctctccc 780  
 ttcgatgggc ctaacctgcc gacgctgaga cagcgggtgc tggagggccg cttccgc 840  
 cccttcttca tgtctcaaga ctgtgagagc ctgatccgc gcatgctggt ggtggacccc 900  
 gccaggcgc 960  
 tcaccatcgcc ccaatccggc cagcaccggc ggtatgcggc tgagccctgc

ttgccgggac ccgcctgccc cgcccttcgc gcacacagct acaccccaa cctggcgac 1020  
 tacgatgagc aggcgctggg tatcatgcag accctggcg tggaccggca gaggacggtg 1080  
 gagtcactgc aaaacagcag ctataaccac tttgtcgcca tttattacct cctcctttag 1140  
 cggtcaagg agtatcgaa tgcccagtgc gcccggcccg ggcctgcccag gcagccgcgg 1200  
 cctcgagct cggacccatcag tggtttggag gtgcctcagg aaggctttc caccgaccct 1260  
 ttccgacctg cctgtcgatc cccgcagccg cagacccatgg tgcagtccgt cctccaggcc 1320  
 gagatggact gtgagctcca gagctcgctg cagtggccct tttttttccc ggtggatgcc 1380  
 agctgcagcg gagtgttccg gccccggccc gtgtccccaa gcagcctgct ggacacagcc 1440  
 atcagtgagg aggccaggca gggggccggc ctagaggagg agcaggacac gcaggagtc 1500  
 ctgcccagca gcacggcccg gaggcacacc ctggccgagg tctccacccg cctctcccc 1560  
 ctcaccgcgc catgtatagt cgtctccccc tccaccacgg caagtccgtc agagggaaacc 1620  
 agctctgaca gttgtctgac cttctctgcg agcaaaaagcc ccgcggggct cagttggcacc 1680  
 ccggccactc aggggctgct gggcgctgc tccccgggtca ggctggcctc gcccttcctg 1740  
 ggttcgcagt cccgcacccccc agtgcgtcg gctcaggggg gcttgggagg agctgttctg 1800  
 ctcctgtca gcttccagga gggacggcg ggcgtcgacca aagggtttct gggactgaac 1920  
 aaggccttc ggcagcagct gaggaagacc acgcggacca ccagccgggc cagcaggggc 1980  
 aaaatcaagg ggctggctcg ccaggtgtgc caggtccctg tgcacggcg cgcaaggccg 2040  
 ggcctgagcc cttccacgc ccctgcacag agcccaggcc agcagaggct gtcaggtta 2100  
 agccgggagg gctggagcct gctggaggag gtgctagagc cccagccggc ccctggcccg 2160  
 cagcaccacc cggccgcgtc accccggctgc tcccaggccc tccccagcac cctccctcact 2220  
 tttgtatcg cccctgtga tggccctggg gctgccccgc ggcgtcccc ggtggctca 2280  
 tcggggctcc cgctgtgcc gcccccaactc ctgcagaccc ggcgtccccc ggtggcctca 2340  
 gcccgcgcgc tcctggacac acacctgcac attggcaccg gccccaccgc ctcctccct 2400  
 gtccccccac cacgcctggc caggctggcc ccagggtgtg agcccccgg gctgtcgac 2460  
 ggggactgtg agatggagga cctgtatgcct tgctccctag tgaggggcagc ctcctccct 2520  
 gtgaagacat ctggcctcaa agccaagaac tttctagaag cgaaataaag aatacgtag 2580  
 gtgtttggc ttttttagttt atttttgttt tatttttttc ttgcactgag tgacctaacc 2640  
 ttttagtagg gactggaaac ttttaggaaga aagataattt aggggcgtgt ctggggcg 2700  
 gggcaggagg ggagcggggt ggagggaaaca cgtgcagtgc cgtgggtgtg ggatctcg 2760  
 ccctctctct gggttcgtcg tggttgagat gattacctcg gacgtctacg gaaacgagcg 2820  
 ggcgcattgt tgtccgttg tgtgtgtgtg tgtgtgtgt actatccatt tcttttagtca acgctctcca cttcctgatt tctgctttaa ggaaaactgt 2940  
 gaactttctg cttcatgtat cagtttaaa gcagcccagg caaagatcat ctacagattc 3000  
 taggaattct ctccccctgaa atcaaaacct ggaagacttt ttttttttat ttttagttgag 3060  
 aagtttcata aactgctcaa ggattagttt tccaggactc tgccggagaa cggcaggaag 3120  
 aacctcagag agggcagagg tgacttcaaa gtgctgggaa ctccgtccctg agggtcactt 3180  
 gcccctgagc ccctgcgtgc cttgcggaa gcccagaagc ttcttcctgc tgacacccccc 3240  
 gttccgtcg ctgctgacgt ttatgcattt catgtgggg tccaacaaga acacctgact 3300  
 tgggtgaagt tgtcaatat tggaggctga ctgtagggtc gggcagctgg gagacaggct 3360  
 catggctcat ggctcatggc tcagggcggt gcctgcccgc ccccccaccc tttgtttttt aagtgagagg ttttaggtcag 3420  
 aggcttttg ggtttgttc aaggaaggta ttttttttaag tttttttaaag caaatccgt atatgtatct acatgggaga 3480  
 caggtagaca ctacttattt gttacatttt gtactacacg tttgtttcc aggtttcagc 3540  
 ttccctcgct cctgttgcata agaagcgtcc ctgtcagcac aggtgtcat tgaggaagg 3600  
 gccccaggcc cttcgctccc tcagcactgg ggtggaggcg gcaggaagg gcccgcctta 3660  
 cctggcaggt ctgggcgcac ctttagcagg tggactccgt ggggctccac cagccagaag 3720  
 ctttggaaag gcaacgaagg caatgctgct ccctgagtcc agtccccggc cccaaaccc 3780  
 gcccagggtc cttcagctac ttggccttct taaaccctgc agtgttaaac agaggcattt 3840  
 agaaaggggaa aaggcgggta tttttaaaag ccaaagattt acccagttac ttgagggtag 3900  
 ggaggcgggc ccagtgcagg aggctgcatt cctggcctgc tgggtccac cggggctgt 4020  
 gcctgtggcg ggccgcagg aagctggctg ccccccattcc tgctgtgtc gctgtgtcg 4080  
 ctctgtggct gtttcaaga ctggcgaaa ggctgtccgg agggcagacc aggtgcctt 4140  
 ccgcagagaa aacaccaaag tctcctgttc gctcataaaag aagtttttgg gatggagag 4200  
 aatccagacc atcttggggc agccaggccc ttgccttcat ttttacagag gtagcacaat 4260  
 tgattccaac acaaaaactcc ttcccccttt taaaatgatt tctgttctaa tgccatagat 4320  
 caaaggcctc agaaaccatt gtgttttcc tctttgaagc aatgacaagc actttactt 4380  
 cacggtggtt tttgtttttt ctattgtcg tggAACCTCT tttggaggac gttaaaggcg 4440  
 ttttttactt gttttttaa gagtggtgtga tttgtgtttt gtagatttct tgacagtgt 4500  
 gtaatacaga cggcaatgca atagcctatt taaagacact acgtgatctg attgagatgt 4560  
 acatagtttt tttttttacc ataactgaat tttttatct cttatgttaa catgagaaat 4620  
 gtatgccaaa tgatttagtt atgtatgttt tttaatttaa tatttaataa aaatatttgg 4680  
 aaggaaaaaa aaaaaaaaaa 4699

<210> 17  
<211> 2649  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<223> Incyte ID No: 387382CB1

<400> 17

```

gcacggaagt ctcgcgatgt ttgcgtttga acctcttggc ggggccggc catggcggct 60
tcgctcccg gacctggag ccggctttc cacacatag gggctgcgga cggcaggaga 120
cagcggccggc cgggcccggg agccgcgcag tggttcccgcc cgccaggacc gaggcggttc 180
ttcaaacagca gcccgcagcag cgacgcgcagc atcggcgacc cctgcgcagtc cgacgatcct 240
gacgatcccg acgaccccgaa cttcccccggc agccccggtaa ggcggccggc aagggtgtccc 300
ggcggccggag tgcccaagga ccggcccggc ctgaccgtga ccccaaagcg ctggaagctg 360
cgagctcgcc caaggctaacc cgtgacccca agacgcctgg ggctgcgagc tcggcccccgg 420
cagaagtgcgca gcacaccctg tggcccgctc cgacttccgc cttcccccag ccgcgactcc 480
ggccgcctca gcccggaccc cagcgtgtgc ggccagccca gggacggcga cgagctgggc 540
atcagtgcct ccctgttcag ctctctggcc tcgcccgtcc cgggtcccc aacgccaagg 600
gacagtgtca tctcgatcggt cacctccggcc tgcgtgggtt cagcctcagc cgtcccgagc 660
ggcctccacc tcccagaagtg ctcctggac cgagcatctc tcccctgctc ccaggaggaa 720
gcgacaggag gagccaagga caccaggatg gtccacccaaa cccgcgcag cctcaggtca 780
gttctctttg gccttatgaa ctcaggaacc cctgaggatt ctgagttcg ggcagatggg 840
aagaatatga gagagtctgt ctgtaaaagg aaactgggtt tggaaatgg accagagggt 900
ccaggtctgt caagcacagg caagaggagg gccacaggcc aggactcttgc tcaagagaga 960
gggcttcaag aggccgtccg gagagagcat caggaggcca gtgttcccaa gggccgcatt 1020
gtgccaaggaa gaatagacag gctggagaga actagatcaa gccggaagag caaacatcag 1080
gaggcaacgg aaacctcttcc cttccattcc caccgcctta aaaagggcca aaagctggga 1140
aaagattcgt tccccaccca ggacctgact ctttacaga atgcctgctt ttggacccaa 1200
accagggtt cttcagttt ccacaagaag aaaattgtga ctgatgtgtc agaggtctgc 1260
agcatctata ccactgccac ttctctcttcc ggatccctcc tatcagaatg ttcaaacccgg 1320
cctgtcatga acagaacaag tggtgctccg tcctcttggc actcctcctc tatgtattt 1380
ctaagccct taaacactct aagtatttca aacaaaaagg catctgatgc tgaaaagggtt 1440
tatggggaat gcagtcagaa gggtcctgtc cccttttagcc attgccttcc cacagaaaaa 1500
ctgcaacgct gtgagaagat tggggaaagggt gtgttggcg aagtgttca aacaattgct 1560
gatcacacac ccgtagccat aaaaatcatt gctattgaag gaccagattt agtcaatgg 1620
tcccatcaga aaaccttga gaaaatcctg ccagagatca tcatctccaa agagttgagc 1680
ctcttatccg gtgaagtgtt caaccgcaca gaaggcttta tcgggctgaa ctcagtgcac 1740
tgtgtccagg gatcttaccc tcccttgctc ctcaaaggctt gggatcacta taattcaacc 1800
aaaggctctg caaatgaccg gcctgatttt tttaaagacg accagcttcc cattgtgtc 1860
gaatttgagt ttggaggat tgacttagag caaatgcgaa ccaagttgtc ttcccttggct 1920
actgcaaaga gcattctaca ccagctcaca gcctccctcg cagtggcaga ggcacactg 1980
cgctttgagc accgagactt acactggggg aacgtgtct taaagaaaac cagcctcaaa 2040
aaactccact acaccctcaa tggaaagagc agcactatcc ccagctgtgg gttgcaagt 2100
agcatcattt actacaccct gtcgcgttcc gaacgggtt ggttgggtt tttctgtgac 2160
gtttccatgg atgaggaccc gtttaccggt gacgggtactt accagtttga catctacagg 2220
ctcatgaaga aggagaataa caaccgcgtgg ggtgaatatc accctttagt taatgtgtc 2280
tggttacatt acctgacaga caagatgttcc aaacaaatga cttcaagac taaatgtaa 2340
actcctgccca tgaagcaaataa taagaaaaaa atccaggagt tccacaggac aatgtgaac 2400
ttcagctctg ccactgactt gctctggccag cacagtctgt ttaagtaagc taaatgtatc 2460
ttactggccc gaaatgagag gagactggtc ttgaaggctt tgggtgtt tcaacctcca 2520
tccccacagg agggtggac tcccattctc acagggttcc agtcagctt ttcaacaaga 2580
attttgttcc aaatttggaaac ctgaatattt gttgaaatgt taaaatttgc tggtacaaat 2640
gttctgttcc 2649

```

<210> 18  
<211> 2526  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature

&lt;223&gt; Incyte ID No: 5850786CB1

&lt;400&gt; 18

tgcgccccgg gagaccggga acatggcgct gggagcgctg tagcagctga gaaggggctg 60  
 aggcaccgcc gcttcgctga cagccggcca ccagaaatac actgtggta aatgcttcca 120  
 cctcttgcta aaatgaacac tgagaaaaaa tgaagaagac tgacaagcac cagcgaaaag 180  
 ttgcagaata gaaatagcca cactcctctg gagtctttaa ttcatccaca gccatcatat 240  
 aaaggtttg gcatcatgtt tggaaagaaa aagaaaaaga ttgaaatatc tggcccgtcc 300  
 aacttgaac acagggttca tactgggtt gatccacaag agcagaagtt taccggcctt 360  
 ccccagcagt ggcacagcct gtagcagat acggccaaca gcccaaagcc tatggtgac 420  
 cttcatgca tcacacccat ccagctggct cctatgaagaa caatcgtag aggaaacaaa 480  
 ccctgcaagg aaacctccat caacggcctg ctagaggatt ttgacaacat ctcggtgact 540  
 cgctccaact ccctaaggaa agaaagccca cccacccag atcaggagc ctccagccac 600  
 ggtccaggcc acgcggaaga aatggcttc atcaccttct cccagtattc cagcaatcc 660  
 gatactactg ctgactacac gaccgaaaag tacagggaga agagtctcta tggagatgat 720  
 ctggatccgt attatagagg cagccacgca gccaagcaaa atgggcacgt aatgaaaatg 780  
 aagcacgggg aggctacta ttctgaggtg aagccttga aatccgattt tgccagattt 840  
 tctgccgatt atcactcaca tttggactca ctgagcaaac caagtgaata cagtgaccc 900  
 aagtgggagt atcagagagc ctcgagtagc tcccctctgg attattcatt ccaattcaca 960  
 cttctagaa ctgcagggac cagcgggtgc tccaaggaga gcctggcgtt cagtgaaagt 1020  
 gaatggggac ccagcctgga tgactatgac aggaggccaa agtcttcgtt cctgaatcag 1080  
 acaagccctc agcccaccat gcggcagagg tccaggtcag gctcgggact ccaggaaccc 1140  
 atgatgccat ttggagcaag tgcattaaa acccatcccc aaggacactc ctacaactcc 1200  
 tacacctacc ctcgcttgcg cgagcccaca atgtgcattt caaaggtgga ttacgatcga 1260  
 gcacagatgg tcctcagccc tccactgtca gggcttgaca cttacccag gggccctgcc 1320  
 aaactaccc aaagtcaaag caaatcgggc tattctcaa gcagtcacca gtacccgtct 1380  
 gggtaccaca aagccaccc ttaccatcac ccctccctgc agagcagttc gcagtcacatc 1440  
 tccacggctt cctacctgag ctccttcagc ctctcatcca gcacccatccc gccgcccagc 1500  
 tgggctcctt cctccgacca gcagccctcc aggggttccc atgaacagtt tcggcggcc 1560  
 ctgcagctgg tggcagccc aggagacccc agggaaatact tggccaaactt tatcaaaatc 1620  
 gggaaaggct caaccggcat cgtatgcac gcccggaga aacacacagg gaaacaagtt 1680  
 gcagtgaaga aaatggaccc ccggaaagcaa cagagacgag aactgcttt caatgaggc 1740  
 gtgatcatgc gggattacca ccatgacaaat gtggttgaca tgtacagcag ctacccgtt 1800  
 ggcgatgagc tctgggttgtt catggagttt ctggatgtt gtccttgac agacattgt 1860  
 actcacacca gaatgaatga agaacagata gctactgtct gcctgtcagt tctgagagct 1920  
 ctctcctacc ttcataacca aggagtgtt cacagggaca taaaaagtga ctccatcctc 1980  
 ctgacaagcg atggccgggt ttcttcagtg ctccgggat tcctagactt gatgttgggt 2040  
 agggagccct ctcagagagc aacagccca gaaactccctcg gacatccatt cttaaaacta 2100  
 gcaggtccac cgtcttcgtt cgtcccccctc atgagacaaat acaggcatca ctgagcagag 2160  
 gattcgtgtt ggtggcaaaag ctagatgagg acatgagaat aattcaggag aacaaaagga 2220  
 aacacagaac atgcaaaagg cctgtgcatt ctagaccagc caattgggtt gacagcgtga 2280  
 tgaccggcag ggttcaacag accagggcat cttcttggtt cttaaacagg catctctcca 2340  
 ctgacagccg gtgtggtcac ttggagcacg gctttataa gtcattatta tattttcag 2400  
 cccttcatcc agcaaatcag aaggactcag tacaaactcc gttatgatat atcctagcca 2460  
 catgcagggc aacatgttagg attttctata ttgaaagaat actttctgg caaaaaaaaa 2520  
 aaaagg 2526

&lt;210&gt; 19

&lt;211&gt; 2178

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Incyte ID No: 7182343CB1

&lt;400&gt; 19

gtggcgcagc cccctgcctg ggtgtccagg ctgtccggc tggggctggg ggagctagag 60  
 gcccacggggg agttggggga gctcctcagg gccccctca cctgcctcc ctgtgcaccc 120  
 ccgtctctct gcagatgaca ttggccagg gaaggagagc ctggggcccg acagtcctc 180  
 tgggggtcaa gaggaccccg ccagccagca gtgggcacga ccgcgttca cacagccctc 240  
 caagatgagg cgccgggtga tcgcacggcc cgtgggttagc tcctgtgcggc tcaagtgcgt 300  
 ggccagcggg caccctcgcc ccgacatcac gtggatgaag gacgaccagg cttgacgcg 360

cccagaggcc gctgagccca ggaagaagaa gtggacactg agcctgaaga acctgcggcc 420  
 ggaggacagc ggcaaataca cctgccgcgt gtcgaaccgc gcggggccca tcaacgccac 480  
 ctacaagggtg gatgtgatcc agcggacccg ttccaagccc gtgctcacag gcacgcaccc 540  
 cgtgaacacg acggtgact tcgggggac cacgtcctc cagtgcagg tgccgacgca 600  
 cgtgaagccg gtgatccagt ggctgaagcg cgtggagta ggcgccgagg gccgccacaa 660  
 ctccaccatc gatgtggcg gccagaagtt tgtggctg cccacgggtg acgtgtggtc 720  
 gcggccccac ggctcctacc tcaataagct gctcatcacc cgtcccccc aggacgatgc 780  
 gggcatgtac atctgcctt ggcacacac catggctac agcttccgca ggccttcct 840  
 caccgtgtc ccagacccaa aaccgccagg gccacctgtg gcctcctcg cctggccac 900  
 tagcctgccg tggccctgg tcatcggcat cccagccggc gctgtttca tcctggcac 960  
 cctgctcctg tggcttgcc aggcccagaa gaagccgtgc accccccgcg ctgcccctcc 1020  
 cctgcctggg caccgcccgc cggggacggc cggcgaccgc agcggagaca aggacattcc 1080  
 ctcgttgcc gcctcagcg ctggccctgg tgtgggctg tgtgaggagc atgggtctcc 1140  
 ggcagcccc cagcaattac tggcccccagg cccagttgtc ggccctaagt tgtaccccaa 1200  
 actctacaca gacatccaca cacacacaca cacacactct cacacacact cacacgtgga 1260  
 gggcaaggtc caccagcaca tccactatca gtgcttagac gcaccgtatc tgcagtggc 1320  
 acgggggggc cggccagaca ggcagactgg gaggatggag gacggagctg cagacgaagg 1380  
 caggggaccc atggcgagga gaaatggca gcacccagg cagtctgtgt gtgaggcata 1440  
 gcccctggac acacacacac agacacacac actacctgga tgcatgtatg cacacacatg 1500  
 cgccgcacacg tgctccctga aggacacacgt acgcacacac gcacatgcac agatatgccc 1560  
 cctggcaca cagataagct gcccaaattgc acgcacacac acagagacat gccagaacat 1620  
 acaaggacat gctgcctgaa catacacacg cacacccatg cgcagatgtg ctgcctggac 1680  
 acacacacac acacggatat gctgtctgga cgcacacacg tgcagatatg gtatccggac 1740  
 acacacgtgc acagatatgc tgcctggaca cacagataat gctgccttga cacacacatg 1800  
 cacggatatt gcctggacac acacacacac acgcgtgcac agatatgtc tctggacagg 1860  
 cacacacatg cagatatgtc gcctggacac acacttccag acacacgtgc acaggcgcag 1920  
 atatgctgcc tggacacacg cagatatgtc gtctagtcac acacacacgc agacatgtc 1980  
 tccggacaca cacacgcac cacagatatg ctgtccggac acacacacgc acgcagatat 2040  
 gctgcctgga cacacacaca gataatgtc cctcaacact cacacacgtg cagatattgc 2100  
 ctggacacac acatgtgcac agatatgtc tctggacatg cacacacgtg cagatatgtc 2160  
 gtccggatac acacgcac 2178

<210> 20  
 <211> 2777  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7264345CB1

<400> 20

ctcagcagag ctaccagctg ccctgttggc ttgcgtggc ggatcgtcct cctggccccc 60  
 ccaaacaggc ggggggagcg gccccgactg tggggccatg gcagtagtct cctcggtcgc 120  
 cggccgcgt agcctagctg agtcgcggc ttctgcgcta ggggctccca ccgcctccgc 180  
 aggctaagga gccgctgcca ccaacgagct gtgagggtt ctagtctccc tcttgccgc 240  
 cgtctccccc tcttgccgc gcaggcaccc ctctggctgc tcagtcctgc ctcagtgtca 300  
 aaccagaaga gaagtaaaaat tcaacaaaaa tttatgtgtc gagttccttc ttaaaaagaag 360  
 aaaaaagtga ttattnagac tatggatcgg agcaaacgga attcaattgc aggatttcct 420  
 ccacgtgtgg agcgtcttga agagttgaa ggagggtgt gaggagaagg aaatgtgagc 480  
 caggtggaa gagttggcc atcttcgtat cgagctctt taagtgcctt ttccagactg 540  
 acgcgtttgg atgatttcac ctgtgaaaaa atagggtctg gcttctttc tgaagtgttc 600  
 aaggtacgac accgagctc tggtcaggtg atggctctt agatgaacac attgagcagt 660  
 aaccgggcaa acatgctgaa agaagtacag ctcataata gactctccca tcccaacatc 720  
 cttaggttca tgggtgtatg ttttcataaa ggacaattgc atgcacttac agatataatc 780  
 aactccggaa acctggaaaca gttgcttagac agtaacctgc atttgcctt gactgtgagg 840  
 gtaaaaactgg cctatgacat agcagtggc ctcagctacc ttcacttcaa aggattttt 900  
 catcgccacc tcacatctaa gaactgcctg ataaagaggg atgagaatgg ttactctgca 960  
 gtggtagctg actttggcct ggctgagaag atccccatg tcagcatggg gagtgagaag 1020  
 ctggccgtgg tgggttcccc attctggatg gcacctgagg ttctccgaga tgagccctat 1080  
 aatgaaaaagg cagatgtgtt ctcttatgtt atcatcctct gcgagatcat cgcccgatc 1140  
 caggccgatc cggactatct tccccgcaca gagaattcg ggctggacta tgcgtttc 1200  
 cagcacatgg tggagactg tccccagat tttctgcaac ttacttcaa ctgctgtaac 1260

atggatccca aactgcgccc atctttgtg gagattggga agaccctgga ggaaattctg 1320  
 agccgcctac aggaagaaga gcaggagagg gataggaagc tgcagcccac agccagggga 1380  
 ctcttggaga aagcacctgg ggtgaagcga ctaagctcac tggatgacaa gatccccac 1440  
 aagtaccat gcccaagacg taccatctgg ctgtctcgaa gccagtcaga tatctttcc 1500  
 cgtaagcccc cacgtacagt gagtgtcttg gaccatact accggccacg agatgggtct 1560  
 gcccgcaccc ccaaagtcaa cccttttagt gctgccagg acctcatggg gggcaagatc 1620  
 aagtttttgc acctgcccag caagtctgtc atctctctgg tatttgacct gnatgaccca 1680  
 gggcccgaa ctatgcccct ggctgactgg caggagcccc tggcccccacc tattcgccgg 1740  
 tggcgttccct tgcctgggtc gcctgagttc ttgcattcaag aggctgtcc atttgtggc 1800  
 cggyaagaat cgctatctga tggggcccca ccacgcctaa gtatgtctaa gtacagagtt 1860  
 aaagagatcc caccattccg ggcattctgcc ctaccagctg ctcaagccca tgaggctatg 1920  
 gactgctcca ttctccagga agaaaatggt tttgggtcca ggccccaggg gaccagtcca 1980  
 tgccctgcgg gtgcttctga ggagatggag gtagaagaaa ggccagcagg ctcaactcca 2040  
 gccaccttct ccacctcagg cataggctg caaaccagg gaaagcagga tgggtgaggg 2100  
 gtttagtcc ctgcctcacc ttggggatgg accttcagct gaaaccatat ggccccctag 2160  
 gtgcacagcc ttgattcttc cctggagcct acagagcagg caggctaggc caagccaggc 2220  
 tcaacttctg ggctcccagt gccattggc tgtgtatgac gggaggcagg agttagaggc 2280  
 cttcttagtt agggccaaca gctgatacca agcctctgaa atccagcaag gaggtctgcc 2340  
 tcccaccaga ccctctccag tgtacttccc cagataggac cagaggatgt ctatgtctag 2400  
 gctgagctgg caggcagcta ttaccccggt tctttccca ccccaaggctt gtctcttgcc 2460  
 ttttcttggg gcatataagc tactgagtgg aacatggagc tgatcaagag gccgtaatgg 2520  
 tcatggctgt ttcccagacc tgaatattgg ggtgcttctt gccagtattc taagacattt 2580  
 gagtaattgc tgtttgcact tactgcatgg tcagaccacg tcactacatt tctatgcaag 2640  
 gggacagcaa ggcagcgtgg tggcatggc tcttagctaa cctattcaaa gaccctttcc 2700  
 tggattaa tctatttca tatttataaa ggagtctaa tgttctgccc cataagactt 2760  
 tcaaccttgc ggttggg 2777

<210> 21  
 <211> 972  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 7472328CB1

<400> 21  
 atgcgagcca cccctttggc tgcttctgcg ggttccctgt ccaggaagaa gcgggtggag 60  
 ttggatgaca acttagatac cgagcgtccc gtccagaaac gagctcgaag tggggccca 120  
 cccagactgc ccccctgcct gttgccctg agcccaccta ctgctccaga tcgtgcaact 180  
 gctgtggcca ctgcctcccg tcttggccc tatgtcctcc tggagcccgaa ggagggcggg 240  
 cgggcctacc aggccctgca ctgcctaca ggcactgagt atacctgcaa ggtgtacccc 300  
 gtccaggaag ccctggccgt gctggAACCC tatgcgcggc tgcccccgga caagcatgtg 360  
 gcttggccca ctgagggcct ggctggtacc caacttctct acgcctttt cacttggacc 420  
 catggggaca tgcacaggct gatagggcac acgcctgtg cgcaactgtga ccagacacgg 480  
 aagaagctgg tgctggagaa cctggaggac tcctgcgtgc tgactggcc agatgattcc 540  
 ctgtgggaca agcacgcgtg cccagctac gtgggacctg agataactcg ctcacggcc 600  
 tcataactcgg gcaaggcagc caatgtctgg agcctggcg tggcgctt caccatgctg 660  
 gccggccact accccttcca ggactcggag cctgtcctgc tcttcggcaa gatccgccc 720  
 ggggcctacg cttgcctgc aggccctcg gccctgccc gctgtctgg tgcgtgcctc 780  
 ctgcgtcggg agccagctga acggctcaca gccacaggca tcctcctgca cccctggctg 840  
 cgacaggacc cgatgcctt agccccaaacc cgatcccattc tctggaggc tgcccaggtg 900  
 gtccctgatg gactgggct ggacgaagcc agggaaagagg agggagacag agaagtggtt 960  
 ctgtatggct ag 972

<210> 22  
 <211> 2538  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <223> Incyte ID No: 1471412CB1

&lt;400&gt; 22

ggggccgtgg ctccgagctg cctagaaagc ggctccaggg agccagggca ccgtgagcct 60  
ggtggttggc agctggagcc acgtcgagg gggaaagtgtc gcagcattct ctgcaggcat 120  
cacagacctg aggcaagtggc ctccggaggg cactggacag aaacagccat ccaagtggct 180  
gagtggaggg accctgctca agtgcagctg cagtggccgg gtttccctc aggttagggat 240  
cggggcgccc tgtcgcccgc agccacgtgt ggcgtccgg acagtcagca gagtcaggg 300  
tgccggcacc aggaaagggg ggcagggga actcccgcgg gcctcgctt tgcaaacttc 360  
tcgcctggc aggaggcggt cgtggaaaag aagggtgaaag agcgagctt ttggaactgt 420  
gcacgggaca gattggacgc acaccctcg ggaggcgcga aggcatggaa aatttgaagc 480  
atattatcac ccttggccag gtcatccaca aacggtgtga agagatgaaa tactgcaaga 540  
aacagtgccg ggcctgggc caccgcgtcc tcggcctgat caagcctctg gagatgctcc 600  
aggaccaagg aaagaggagc gtgcctctg agaagttAAC cacagccatg aaccgcttca 660  
aggctgccct ggaggaggct aatggggaga tagaaaagtt cagcaataga tccaatatct 720  
gcaggttct aacagcaagc caggaaaaaa tactcttcaa ggacgtgaac aggaagctga 780  
gtgatgtctg gaaggagctc tcgctttac ttcaggttga gcaacgcattg cctgtttcac 840  
ccataaggca aggagcgtcc tgggcacagg aagatcagca ggatgcagac gaagacaggc 900  
gagctttca gatgctaaga agagataatg aaaaaataga agtttcaactg agacgattag 960  
aaatcaacat gaaagaaaatc aaggaaactt tgaggcagta tttaccacca aaatgcattgc 1020  
aggagatccc gcaagagcaa atcaaggaga tcaagaagga gcagctttca ggatccccgt 1080  
ggattctgct aaggaaaaat gaagtcagca cactttataa aggagaatac cacagagctc 1140  
cagtggccat aaaagtattc aaaaaactcc aggctggcag cattgcaata gtgaggcaga 1200  
ctttcaataa ggagatcaaa accatgaaga aattcgaatc tcccaacatc ctgcgttat 1260  
ttgggatttgc cattgatgaa acagtgactc cgcctcaatt ctccattgtc atggagttact 1320  
gtgaactcg gaccctgagg gagctttgg ataggaaaa agacccatc cttggcaagc 1380  
gcatggtcct agtccctggg gcagcccgag gcctataacc gctacaccat tcagaagcac 1440  
ctgaactcca cggaaaaatc agaagctcaa acttcctggt aactcaaggc taccaaatgt 1500  
agcttgcagg atttgagttg aggaaaacac agacttccat gagtttggaa actacgagag 1560  
aaaagacaga cagagtcaaa tctacagcat atctctcacc tcaggaactg gaagatgtat 1620  
tttatcaata tgatgtaaag tctgaaatat acagctttgg aatcgcttc tggaaaatcg 1680  
ccactggaga tatcccgttt caaggctgtt attctgagaa gatccgcaag ctggggctg 1740  
tgaagcggca gcaggagcca ctgggtgaag actgccttc agagctgcgg gagatcattg 1800  
atgagtgccg ggcccatgat ccctctgtc ggccctctgt gnatgaaatc taaaagaaaac 1860  
tctccacatt ttcttaatgtatgttcaaaa tctaaaccaa ggagtctctg gacaagaagc 1920  
tgggagaggg acaaacttgg catctctctc tctcatatcc ttcggcattt ggttatctat 1980  
gggagcaagg agtgggcacg cttctctgtt acaaataaaaa acacatccca aacgattcca gtcatacagg 2040  
acacatccca ctccaaatga tatttccaaa aacataacctc tgacagtaac tttgatagat 2100  
ggtttgcattt atgtatctt ctgggtatcc acacctcttgc gcaatgaaat ttgcagctcc 2160  
tcccttcattt aatgaagtc tctttccccca ccatttgaat ctgggtggc actgtgactt 2220  
gatttgcattt atagaatgttga gaagaagtgttgc gttccaaagcc taggtttcaat 2280  
gaggccttat aatgtctgt tggaaacctt cccagccatg aacatgttgc gtgagcatgc 2340  
tggagaatgttga gagaccacat gaagcagaaa catgcattcc tagctgaagt catactagcc 2400  
caaccaacat ggcagctaac acatgaatgttga ggcataatcaa gaccagaaga accactcaag 2460  
cagatcccag cccaaatttgc ccattcacac aatcaggagc taaataaattt actgttgcattt 2520  
aaaaaaaaaaaaaaa aaaaaaaaaaaaaaaa 2538